KU LEUVEN

FACULTEIT PSYCHOLOGIE EN
PEDAGOGISCHE WETENSCHAPPEN
LABO VOOR EXPERIMENTELE PSYCHOLOGIE

Some implicatures are conversational, *but* others are conventional: a developmental implicature study

Proefschrift aangeboden tot het verkrijgen van de
graad van Doctor in de Psychologie
door **Leen Janssens**
o.l.v. Prof. dr. Walter Schaeken (promotor)

2014

Leen Janssens. ***Some* implicatures are conversational, *but* others are conventional: a developmental implicature study**
Dissertation submitted in order to obtain the degree of Doctor in Psychology, 2014.
Supervisor Prof. dr. Walter Schaeken

In communication, people often intend to convey more than the words they literally utter. The literal, semantic meaning of a sentence can be enriched by its implicit, pragmatic meaning. The term *implicature* refers to an interpretation of what a speaker didn't explicitly say but intended to say. Two broad categories of implicatures can be distinguished. On the one hand, there are conversational implicatures of which the scalar implicature is the most widely investigated subtype. On the other hand, there are conventional implicatures which have rarely been experimentally investigated. This dissertation discusses both types of implicature from a developmental point of view.

In scalar implicature research, it has often been concluded that adults are more pragmatically competent than children. This was interpreted as indirect evidence that inferring a scalar implicature is cognitively effortful and requires working memory. Other evidence showing that working memory is involved in scalar implicature processing was presented among others by Bott and Noveck (2004) and De Neys and Schaeken (2007). However, Katsos and Bishop (2011) showed that certain task characteristics can conceal children's pragmatic competence and therefore lead to the wrongly drawn conclusion that children are pragmatically incompetent. This shows that it is worth investigating the role of age, task characteristics and working memory in implicature research, especially for conventional implicatures of which little experimental information is available.

In Chapter 2 of this dissertation, a scalar implicature study is described in which we look at the effect of age, working memory capacity and task features on processing the scalar implicature from *some*. Age and task characteristics were found to greatly influence the number of pragmatic responses whereas an effect of working memory was absent.

In Chapters 3 to 7, the conventional implicature stemming from *but* –combined with *so* and *nevertheless*- was investigated. Our primary interest was finding out whether the conventional meaning of these instruction words is indeed understood. Secondly, we investigated the effect of age, working memory and task characteristics as well. Our results showed that adults seem to have a pretty good understanding of the conventional meaning of these words whereas children's performance revealed lower competence. In fact, using a three-point scale instead of a binary response format in Chapter 5 seemed to reveal that these sentences are really difficult for children and that they often don't know what they have to answer. With regards to the working memory effect, we found no influence of working memory on conventional implicature processing. This leads to the conclusion that conventional implicature processing happens automatically.

Leen Janssens. ***Sommige*** **implicaturen zijn conversationeel, *maar* anderen zijn conventioneel: een ontwikkelingsstudie naar implicaturen.**
Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Psychologie, 2014.
Promotor Prof. dr. Walter Schaeken

Tijdens het communiceren willen mensen dikwijls meer zeggen dan de woorden die ze letterlijk gebruiken. De impliciete, pragmatische betekenis van een zin kan de letterlijke, semantische betekenis verrijken. *Implicaturen* verwijzen naar interpretaties van wat een spreker niet letterlijk zegt, maar wel bedoelt te zeggen. Twee categorieën implicaturen kunnen onderscheiden worden. Enerzijds zijn er conversationele implicaturen, waarvan de scalaire implicatuur het meest onderzochte subtype is. Anderzijds zijn er conventionele implicaturen die zelden experimenteel onderzocht zijn. Dit proefschrift bespreekt beide types van implicaturen vanuit een ontwikkelingsperspectief.

In onderzoek naar scalaire implicaturen werd er dikwijls geconcludeerd dat volwassenen pragmatisch competenter zijn dan kinderen. Dit werd beschouwd als indirect bewijs dat het cognitief belastend is om een scalaire implicatuur af te leiden. Ook Bott en Noveck (2004) en De Neys en Schaeken (2007) vonden bewijs dat werkgeheugen een rol speelt bij het verwerken van scalaire implicaturen. Echter, Katsos en Bishop (2011) hebben aangetoond dat bepaalde taakkenmerken ervoor kunnen zorgen dat pragmatische competentie bij kinderen afwezig lijkt, terwijl dat niet zo is. Dit kan leiden tot de verkeerde conclusie dat kinderen pragmatisch incompetent zijn. Bijgevolg is het interessant om het effect van leeftijd, taakkenmerken en werkgeheugen na te gaan in onderzoek naar implicaturen, zeker bij conventionele implicaturen waarover weinig experimentele informatie beschikbaar is.

In Hoofdstuk 2 van dit proefschrift wordt er een studie naar scalaire implicaturen beschreven waarin we kijken naar de verwerking van de scalaire implicatuur uit *sommige*. De resultaten toonden aan dat leeftijd en taakkenmerken een zeer grote invloed uitoefenen op het aantal pragmatische antwoorden terwijl een werkgeheugeneffect afwezig bleek.

In de Hoofdstukken 3 tot en met 7 werd de conventionele implicatuur uit *maar* –in combinatie met *dus* en *toch*- onderzocht. In eerste instantie wilden we achterhalen of de conventionele betekenis van deze instructiewoorden inderdaad begrepen wordt. Ten tweede onderzochten we ook hier het effect van leeftijd, werkgeheugen en taakkenmerken. Onze resultaten toonden aan dat volwassenen een redelijk goed begrip hebben van de conventionele betekenis van deze woorden, terwijl kinderen minder competent leken. Meer zelfs, wanneer er een driepuntenschaal gebruikt werd in plaats van een binair antwoordformaat in Hoofdstuk 5, bleken de resultaten aan te geven dat kinderen vaak niet weten wat te antwoorden. Wat betreft het effect van werkgeheugen vonden we geen invloed van werkgeheugen op het verwerken van conventionele implicaturen. Hieruit concluderen we dat het verwerken van conventionele implicaturen automatisch gebeurt.

Dankwoord

Uiteraard zou dit doctoraat nooit tot stand zijn kunnen komen zonder de hulp van enkele fantastische mensen die ik dan ook van harte wil bedanken.

In de eerste plaats wil ik Walter (a.k.a de tofste promotor ter wereld) bedanken voor zijn geweldige begeleiding gedurende de voorbije jaren. Walter, bedankt om mij elke keer opnieuw gerust te stellen als ik weer eens onnodig panikeerde of me onzeker voelde. Je stond altijd klaar met positieve, constructieve feedback en je gaf me altijd het vertrouwen dat alles wel in orde zou komen. Ik heb het altijd heel plezant gevonden om met jou samen te werken.

Ons RAT-groepje is intussen hard uitgedund, maar behalve de huidige RATjes wil ik ook zeker enkele ex-RATjes bedanken. Sara, Aline (en Ilse reken ik daar voor het gemak ook even bij), ik heb altijd een beetje naar jullie opgekeken als naar een grote, wijze zus. Zowel op academisch vlak als op ontspannend vlak heb ik ontzettend hard van jullie goede raad en van jullie gezelschap genoten.

Katrijn en Karolien, ook jullie nemen een zeer belangrijke plaats in dit dankwoord in. Jullie zijn allesbepalend geweest voor de sfeer waarin ik de voorbije jaren heb kunnen werken. Ik kon altijd beroep doen op jullie voor werkgerelateerde vragen, maar bovenal hebben jullie me de voorbije jaren ontzettend goed kunnen entertainen. Grappige verhalen, zotte uitspraken, vreemde google-opzoekingen, en dat gecombineerd met onvergetelijke zee-uitstapjes, sushi-etentjes, kerstfeestjes,… Het heeft mij de voorbije jaren de broodnodige ontspanning bezorgd. Bedankt ladies!

Ook de andere labo-genoten (Nathalie, Pieter, Eef, Bart, Maarten,...en al de rest) wil ik bedanken voor de leuke after-work feestjes, labdrinks, lunches, KUBB-toernooien, BBQ's etc.

Marleen, ook jij verdient een welgemeende dankuwel. Wat zou ons labo zonder jou moeten beginnen?

I would sincerely like to thank the members of my supervisory committee: Napoleon Katsos, Wim De Neys, William Van Belle and Eef Ameel. Thank you for reading my doctoral project and for providing me with very useful feedback. It truly was an honor for me that some of the big names within the implicature research field, who I look up to, agreed to read and comment on my work. A special thanks to Napoleon Katsos, Wim De Neys and William Van Belle who also accepted to be a member of my doctoral jury and to read this dissertation.

Wie eveneens een onnoemelijk belangrijke plaats in dit dankwoord innemen zijn mijn ouders. In de eerste plaats uiteraard bedankt om mij de kans te geven om deze studies en dit doctoraat te kunnen aanvatten. Maar vooral bedankt voor jullie grote steun en

hulp, op elk levensvlak, de voorbije jaren. Jullie zijn zonder twijfel de meest behulpzame mensen die ik ken, niet alleen naar mij toe, maar voor iedereen in jullie omgeving. Bedankt!

Geert, mijn 'big bro', en Jona: Ook jullie wil ik bedanken voor alle ontspannende familie-momenten de voorbije jaren. Ook jullie twee kleine bengeltjes – Milo en Magnus – zijn een bedankje waard omdat hun schattige snoetje zien elke keer opnieuw een grote glimlach op mijn gezicht tovert.

Katrien, mijn echte 'grote wijze' zus, maar ook mijn beste maatje: Bij jou voel ik me altijd het kleine, goed verzorgde zusje. Net als mama en papa sta je altijd en overal voor mij paraat en daarvoor wil ik je ontzettend hard bedanken. Ook Niko, mijn schoonbroer, wil ik bedanken om er altijd te zijn als ik het nodig heb. En jullie meisjes zijn dan misschien nog maar 'vers van de pers', maar ook nu al bezorgen zij mij heel veel energie en een gelukkig gevoel als ik alleen al maar aan hen denk.

Ook al mijn vrienden wil ik bedanken om gedurende alle jaren van mijn doctoraat (en uiteraard ook daarvoor al) te zorgen voor leuke, ontspannende momenten. Een speciale vermelding is hier toch wel op zijn plaats voor mijn vier allerbeste vriendinnetjes – Charlotte, Isabelle, Sarah en Sofie - op wie ik al heel mijn leven kan terugvallen. Bedankt voor alle woensdagavond-etentjes, uitstapjes naar de zee, naar Den Haag, chirofeestjes, shoppingtrips etc.

Tenslotte, last, maar allesbehalve least, Philip: Ik heb ontzettend veel bewondering voor jouw kracht, jouw doorzettingsvermogen en jouw vastberadenheid. Ik wil je ontzettend hard bedanken voor jouw onvoorwaardelijke steun, zelfs in tijden van chronische verbouwingsvermoeidheid. Op het vlak van mijn doctoraat had ik veel aan de raad en hulp van 'Dr. Breugelmans', maar je hebt me ook geleerd om de dingen te leren relativeren. Ik weet niet wat ik zonder jou zou moeten doen. Bedankt voor alles!

Table of Contents

CHAPTER 1 INTRODUCTION AND OUTLINE 1

IMPLICATURES 1
CONVERSATIONAL IMPLICATURES 2
CONVENTIONAL IMPLICATURES 6
OUTLINE OF THE EXPERIMENTAL STUDIES 8
CHAPTER *2* 8
CHAPTER 3 9
CHAPTER 4 10
CHAPTER 5 10
CHAPTER 6 11
CHAPTER 7 11
CHAPTER 8 11
REFERENCES 12

CHAPTER 2 'SOME': EFFECTS OF AGE, TASK, TASK CONTENT AND WORKING MEMORY 15

ABSTRACT 15
INTRODUCTION 16
EXPERIMENT 1 18
METHOD 19
RESULTS 21
DISCUSSION 22
EXPERIMENT 2 23
METHOD 23
RESULTS 24
DISCUSSION 26
EXPERIMENT 3 27
METHOD 27
RESULTS 28
GENERAL DISCUSSION 29
REFERENCES 33

CHAPTER 3 'BUT' HOW DO WE REASON WITH IT: AN EXPERIMENTAL INVESTIGATION OF THE IMPLICATURE STEMMING FROM 'BUT' 35

ABSTRACT 35
INTRODUCTION 36
EXPERIMENT 1 44
METHOD 44

Results 46
Discussion 47
Experiment 2 **49**
Method 49
Results 50
General discussion **52**
References **55**

CHAPTER 4 BUT: DO AGE AND WORKING MEMORY INFLUENCE CONVENTIONAL IMPLICATURE PROCESSING? **69**

Abstract **69**
Introduction **70**
Method **75**
Participants 75
Working Memory Task 75
Implicature Task 76
Procedure 78
Results **78**
General discussion **81**
References **85**

CHAPTER 5 'BUT' HOW DO CHILDREN JUDGE IT ON A SCALE? **87**

Abstract **87**
Introduction **88**
Experiment 1 **93**
Method 93
Results 95
Discussion 96
Experiment 2 **97**
Method 97
Results 97
General Discussion **98**
References **101**

CHAPTER 6 'BUT' LOAD DOESN'T MATTER: THE AUTOMATICITY OF PROCESSING CONVENTIONAL IMPLICATURES **103**

Abstract **103**
Introduction **104**
Experiment **109**
Method 109
Results 111

GENERAL DISCUSSION **116**
REFERENCES **120**

CHAPTER 7 IT'S MAYBE SOMEWHAT DIFFICULT *BUT* I UNDERSTAND IT! **123**

ABSTRACT **123**
INTRODUCTION **124**
EXPERIMENT **130**
METHOD 130
RESULTS AND DISCUSSION 131
GENERAL DISCUSSION **132**
REFERENCES **134**

CHAPTER 8 SUMMARY AND FINAL DISCUSSION **137**

SUMMARY OF THE EXPERIMENTAL FINDINGS **137**
CHAPTER 2 137
CHAPTER 3 139
CHAPTER 4 139
CHAPTER 5 140
CHAPTER 6 140
CHAPTER 7 141
FINAL DISCUSSION **141**
TASK CHARACTERISTICS 141
DEVELOPMENT 144
WORKING MEMORY 149
REFERENCES **157**

List of Tables

CHAPTER 1 INTRODUCTION AND OUTLINE 1

TABLE 1 THE CONVERSATIONAL MAXIMS OF GRICE (1989) 2

CHAPTER 2 'SOME': EFFECTS OF AGE, TASK, TASK CONTENT AND WORKING MEMORY 15

TABLE 1 PERCENTAGES OF RESPECTIVELY CORRECT AND LOGICAL ANSWERS ON THE CONTROL SENTENCES AND CRITICAL SENTENCES OF THE ABT AND TVJT FOR LOW- AND HIGH WM SPAN CHILDREN (EXPERIMENT 2) 25

TABLE 2 PERCENTAGES OF RESPECTIVELY CORRECT AND LOGICAL ANSWERS ON THE CONTROL SENTENCES AND CRITICAL SENTENCES OF THE ABT, TVJT AND WORLD-KNOWLEDGE TVJT FOR LOW- AND HIGH WM SPAN CHILDREN (EXPERIMENT 3) 29

CHAPTER 3 'BUT' HOW DO WE REASON WITH IT: AN EXPERIMENTAL INVESTIGATION OF THE IMPLICATURE STEMMING FROM 'BUT' 35

APPENDIX A: MATERIALS EXPERIMENT 1 (TRANSLATED FROM DUTCH) 57

APPENDIX B: MATERIALS EXPERIMENT 2 (TRANSLATED FROM DUTCH) 62

CHAPTER 4 BUT: DO AGE AND WORKING MEMORY INFLUENCE CONVENTIONAL IMPLICATURE PROCESSING? 69

TABLE 1 EXAMPLE OF EACH SENTENCE TYPE 77

TABLE 2 PERCENTAGES OF APPROPRIATE SO- AND NEVERTHELESS-CONCLUSIONS FOR EACH SENTENCE TYPE 79

TABLE 3 PARAMETER ESTIMATES FOR THE MODEL WITH SENTENCE TYPE AS A PREDICTOR 80

TABLE 4 PAIRWISE COMPARISONS FOR THE DIFFERENT SENTENCE TYPES 81

CHAPTER 5 'BUT' HOW DO CHILDREN JUDGE IT ON A SCALE? 87

TABLE 1 INDICATION OF WHICH ARGUMENT HAS MORE WEIGHT FOR EVERY SENTENCE CATEGORY 92

TABLE 2 PERCENTAGES OF NEUTRAL AND APPROPRIATE (BETWEEN BRACKETS) ANSWERS FOR EACH ARGUMENT CONCLUSION COMBINATION (EXPERIMENT 1) 95

TABLE 3 PERCENTAGES OF NEUTRAL AND APPROPRIATE (BETWEEN BRACKETS) ANSWERS FOR EACH ARGUMENT CONCLUSION COMBINATION (EXPERIMENT 2) 98

TABLE 4 RESULTS OF MANN-WHITNEY U TESTS AND PERCENTAGES FOR THE COMPARISON BETWEEN THE TWO AGE GROUPS 98

CHAPTER 6 'BUT' LOAD DOESN'T MATTER: THE AUTOMATICITY OF PROCESSING CONVENTIONAL IMPLICATURES 103

TABLE 1 PARAMETER ESTIMATES FOR THE FINAL MODEL 113

CHAPTER 7 IT'S MAYBE SOMEWHAT DIFFICULT *BUT* I UNDERSTAND IT! 123

TABLE 1 MEAN FEELING OF UNDERSTANDING SCORE IN THE CONDITIONS WITH 'I UNDERSTAND' IN THE P- OR THE Q-ARGUMENT AND WITH A PERIOD OR BUT AS A CONNECTOR BETWEEN THE ARGUMENTS 132

List of Figures

CHAPTER 3 'BUT' HOW DO WE REASON WITH IT: AN EXPERIMENTAL INVESTIGATION OF THE IMPLICATURE STEMMING FROM 'BUT' **35**

FIGURE 1. **PERCENTAGES OF APPROPRIATE ANSWERS FOR EACH OF THE ARGUMENT COMBINATIONS.** **47**

CHAPTER 6 'BUT' LOAD DOESN'T MATTER: THE AUTOMATICITY OF PROCESSING CONVENTIONAL IMPLICATURES **103**

FIGURE 1**. INTERACTION BETWEEN CONNECTOR AND CONCLUSION TYPE.** **114**

FIGURE 2A, B, C, D**. INTERACTIONS BETWEEN CONCLUSION TYPE AND AXIOLOGICAL VALUE COMBINATION FOR EACH ARGUMENT COMBINATION (A) WEAK-STRONG, (B) STRONG-WEAK, (C) WEAK-WEAK, (D) STRONG-STRONG.** **115**

Chapter 1

Introduction and Outline

This chapter starts with a general introduction, discussing the framework of this dissertation. First, I will explain what implicatures are and then I will focus more closely on each of the two major parts of this dissertation: the two subcategories of conversational and conventional implicatures. I will finish this chapter by providing an overview of the experimental studies discussed in more detail in the next chapters.

Implicatures

Even though every-day-communication seems to happen automatically, verbal communication is a complex phenomenon. The communication process involves a lot more than just the simple encoding and decoding of a message by a messenger and a receiver. Both the semantic aspects, or the literal meaning of a sentence, and the pragmatic aspects, or the implicit meaning the speaker wants to communicate, have to be acknowledged. "Speakers often intend to convey far more than the words they utter and hearers manage to go beyond what speakers have uttered and manage to retrieve the intended interpretation of the utterance" (Papafragou & Musolino, 2003, p. 254). According to Grice (1975), communication is a cooperative enterprise between people. He formulated his cooperative principle as follows:

> *Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.* (Grice, 1975, p.46)

Grice (1989) proposed four relational expectations or maxims that interlocutors should adhere to in order to ensure smooth communication. These maxims and their descriptions are presented in Table 1.

Table 1 *The conversational maxims of Grice (1989)*

Maxim of Quantity	(1) Make your contribution as informative as is required (for the current purposes of the exchange)
	(2) Do not make your contribution more informative than is required
Maxim of Quality	Be truthful
	(1) Do not say what you believe to be false
	(2) Do not say that for which you lack adequate evidence
Maxim of Relation	Be relevant
Maxim of Manner	Be perspicuous
	(1) Avoid obscurity of expression
	(2) Avoid ambiguity
	(3) Be brief (avoid unnecessary prolixity)
	(4) Be ordered

Grice introduced the term 'implicature', being an interpretation of a speaker's intention of what is not being said explicitly. The broad category of implicatures can be subdivided into conversational implicatures and conventional implicatures (Grice, 1989). Horn (2004) defines a conversational implicature as "a component of speaker meaning that constitutes an aspect of what is meant in a speaker's utterance without being part of what is said" (Horn, 2004, p.3). A conventional implicature is defined by Huang (2006) as "a non-truth-conditional inference which is not deductive in any general, natural way from the saying of what is said, but arises solely because of the conventional features attached to particular lexical items and/or linguistic constructions" (Huang, 2006, p.54). The two major parts of this dissertation coincide with these two subcategories of implicatures.

Conversational implicatures

The category of conversational implicatures can be further subdivided into two different types: generalized and particularized conversational implicatures. These two types differ in their need for specific contextual information in order to derive the inferences. Particularized implicatures require contextual information. For example:

(1) Kate: "Are you having dinner with me tonight?"
Steve: "I'm going to a party with some friends."

Kate can infer from Steve's utterance in (1) that he will not join her for dinner even though Steve didn't explicitly say that. This implicature is context-dependent because Steve's answer would not have the same meaning in a different context. On the other hand, generalized conversational implicatures do not require specific contextual information. This type of inference is based on the assumption that the speaker follows the cooperation principle and the Gricean maxims. An example is:

(2) Anna: "I am talking to a man."

The generalized implicature from the utterance in (2) is that Anna is not talking to her husband. According to Grice's maxim of quantity, it can be assumed that the speaker is as informative as possible. Consequently, if Anna would have been talking to her husband, it can be expected that she would have given this more informative information.

A widely investigated example of a generalized implicature is the *scalar implicature*. This type of implicature is inferred from words that can be placed on a scale, according to the amount of information they express, such as <*all, most, many, some*>, <*certain, probable, possible*>, <*three, two, one*>, etc. When a weaker term from such a scale is expressed, it can be inferred that a stronger term from that scale does not apply. For example, even though the utterance *'John has two children'* is logically true if John has three children, the implicature can be inferred that John has exactly two children.

Over the last decades, a whole range of different studies have experimentally investigated scalar implicatures in both (young) children and adults. A considerable number of these studies have concluded that preschool children are often insensitive to scalar implicatures in tasks involving language comprehension (e.g., Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Musolini & Lidz, 2002; Noveck, 2001). In these studies, linguistically competent children seemed to prefer the logical/semantic meaning of scalar terms. This finding led Noveck (2001) to conclude that "younger, albeit competent reasoners, initially treat a relatively weak term logically before becoming aware of its pragmatic potential", and that, in this respect, "children are more logical than adults" (Noveck, 2001, p. 165). This pragmatic delay in children has been typically explained by referring to cognitive capacities. For example, Noveck (2001; also

see Guasti et al., 2005) suggested that inferring scalar implicatures is cognitively effortful, so the fact that children have less available cognitive resources than adults is a plausible explanation for the pragmatic delay in children.

Regarding the cognitive processing costs of inferring scalar implicatures, two different theories make different predictions. According to the default theories (including the neo-Gricean theory; e.g., Levinson, 2000), implicature production happens automatically and only its inhibition demands processing costs. Chierchia et al. (2001) argued in favor of a default theory. They predicted that scalar implicatures only will not be produced when they are embedded in a downward-entailing context (e.g., in the case of negations and question forms), but in all other cases the inference will be produced (automatically). Levinson (2000) argued that scalar implicatures arise because of a Q-heuristic. This heuristic is closely related to Grice's first maxim of quantity and entails that 'what isn't said, isn't the case'. He claims that upon hearing a weak term, the assumption is automatically made that the speaker knows that a stronger term isn't true or that the speaker doesn't have enough information to know whether it's true.

In contrast to the default theories, the contextual theories (e.g., Sperber & Wilson, 1995), suggest that an implicature will only be produced if it is relevant in the context and that this production requires additional processing costs. For example, according to relevance theory, a hearer will always try to reach an interpretation of an utterance with maximal effect and minimal effort. When a logical interpretation leads to a satisfying understanding of an utterance, there is no need to enrich its meaning pragmatically. However, when its meaning has to be pragmatically enriched in order to be relevant in the context, this requires effort.

The child studies discussed above provide indirect evidence in favor of the contextual view. Because children seem to prefer logical over pragmatic interpretations and because their cognitive capacities aren't as developed as adults' capacities, this is indirect evidence that processing scalar implicatures is cognitively effortful. Other studies presented more direct evidence (e.g., Bott & Noveck, 2004; De Neys & Schaeken, 2007; Noveck & Posada, 2003). Bott and Noveck (2004, Experiment 4) manipulated the availability of cognitive resources by varying response time. When participants were limited to only 900ms, the number of pragmatic answers decreased significantly compared to the condition in which there was more time available for answering.

Similarly, De Neys and Schaeken (2007) burdened participants' working memory capacity with a secondary task and found a significant drop in pragmatic answers under cognitive load. Finally, Noveck and Posada (2003) used Evoked Potential techniques and concluded from their study that implicatures emanate from a late-arriving, effort-demanding decision process.

Apart from studies on the pragmatic delay in children, there is a lot of experimental evidence showing that children are aware of the pragmatic potential of scalar expressions and are not incapable to draw scalar inferences. Katsos and Smith (2010), for example, criticized the typically used tasks in most implicature studies. They found evidence in Experiment 1 that young children did not reject underinformative utterances in a binary judgement task, whereas they did penalize these same utterances when given a five-point Likert scale in Experiment 2. This finding is in line with their Pragmatic Tolerance Hypothesis which claims that young children are pragmatically competent comprehenders, but they operate under a principle of tolerance towards pragmatic violations which leads them to not reject underinformative utterances when they are only given a binary choice. Likewise, other studies showed that the used implicature task greatly influences performance and therefore leads to the distorted conclusion that children are pragmatically incompetent. For example, Chierchia et al. (2001) used a Truth Value Judgement Task (TVJT) (Crain & McKee, 1985; Crain & Thornton, 1998) in their first two experiments and a Felicity Judgement Task (FJT) in a third experiment for investigating the scale <*and, or*>. In a TVJT, participants are instructed to judge whether statements are right or wrong. In a FJT, participants are presented with two alternative descriptions of a specific situation and are instructed to indicate which one is the best representation of the given context. Both alternatives have the same truth-value in the context under consideration but differ in appropriateness. Using the FJT increased the number of pragmatic answers to 93%, compared to 50% on the TVJT (Chierchia et al., 2001).

In conclusion, there is abundant experimental literature available on processing conversational implicatures and more specifically on scalar implicatures. The most widely investigated scalar is *some* but others such as *might*, *or* and quantity implicatures have been documented as well. The comparison between different types of scalars has shown that it depends on the type of scale how easily the implicature is derived (e.g., Papafragou & Musolino, 2003).

Most scalar implicature research performed on children has concluded that there is a pragmatic delay in children compared to adults. This would imply that working memory capacity is involved in processing scalar implicatures. However, as Katsos and Bishop (2011) have shown, this delay can be a wrongly interpreted side effect of the used task format.

In the scalar implicature investigations reported in this dissertation, we will focus on these critical aspects that can affect implicature processing in general. First of all, we approach this topic from a developmental viewpoint, investigating children as young as three years up to children of seven years old. Second, we explicitly manipulate the nature of the task as well as the specific task content and look how this affects the number of inferred scalar implicatures. Third, we look at the relation between working memory and scalar implicature processing as this has never been tested in children.

Conventional implicatures

In contrast to conversational implicatures, conventional implicatures have rarely been experimentally investigated. "Conventional implicatures are not derived from the cooperative principle and its component maxims, but are attached by convention to particular lexical items or linguistic constructions" (Huang, 2006, p.56). Unlike conversational implicatures, conventional implicatures are related to the conventional meaning of words, are immediate conclusions from utterances, they cannot be cancelled and they are related to the form of an utterance, not the content. The conventional implicature elaborated in this dissertation is the implicature stemming from *but*. The conventional meaning of *but* in a 'p but q' sentence is that there is a contrast between *p* and *q*. However, this contrast is not explicitly expressed. For example:

(3) She is cute, but she is smart.

(4) She is cute and she is smart.

The utterance in (3) implies that the speaker assumes that being cute and being smart are usually not compatible. If *but* would be replaced by *and*, as in (4), the sentence would still have the same truth-conditions. Both (3) and (4) imply that the subject of this sentence is cute and smart. Therefore, their purely semantic meaning is the same. However, pragmatically, the word *but* adds extra meaning to the sentence, namely that

there is a contrast between the two parts. This additional meaning is not elicited by the connector *and*.

According to Jasinskaja (2012), there are three types of uses of words such as *but*, that are most commonly discussed. The examples below, adopted from Jasinskaja (2012), illustrate these types:

(5) This ring is beautiful, but that one isn't.

(6) This ring is beautiful, but expensive.

(7) This ring is beautiful, but we won't buy it.

The use in (5) is called the *formal contrast* or *semantic opposition* use. In these sentences, *but* highlights the similarities and differences between two propositions. In (6), *but* is used *argumentatively* and indicates that there are two contrastive arguments expressed for the same suggestion (i.e. that we should buy the ring). The third use, exemplified in (7), is the *concessive* or *denial of expectation* use. In these sentences, the inference from the first conjunct is denied in the second conjunct (Jasinskaja, 2012).

In this dissertation, we will discuss *but* in its argumentative use. In the terminology of the Algemene Nederlandse Spraakkunst (ANS; General Dutch Grammar), we will look at *but* as a distancing contrastive connector (Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn, 1997). In this kind of 'p but q' sentence, the second part of the complex speech act gets disconnected from the first part by *but*. However, the first part *p* is still recognized as true (Van Belle & Devroy, 1992). The conventional meaning of *but* cancels the inference from *p* and allocates more weight to the inference from *q*. As in (7), *but* can directly cancel the inference from *p* by expressing the opposite conclusion in *q*, which is called a direct concession in Van Belle and Devroy (1992). However, in this dissertation, we mainly focus on indirect concessions, such as (6), in which the inference from *q* overrules the inference from *p* because of the conventional meaning of *but*. As a consequence, the conclusion arising from (6) is that we will not buy the ring.

The conventional meaning of the words *so* and *nevertheless*, combined with an indirect concession, leads to opposite conclusions. *So* confirms and strengthens the inference from *q*, as shown in (8):

(8) This ring is beautiful, but expensive. So I don't buy it.

In contrast, Van Belle (2003) argues that *nevertheless* (translated from Dutch *toch*) elicits the conclusion from *p.* For example:

(9) This ring is beautiful, but expensive. Nevertheless I buy it.

In conclusion, in contrast to conversational implicatures, not much is known about how these conventional implicatures are understood and processed. In this dissertation, we wanted to draw a parallel between our two lines of research. That's why we were interested in the influence of the same aspects on both conversational and conventional implicatures. Consequently, our conventional implicature studies focus on the role of working memory, task characteristics and development as well.

Outline of the experimental studies

In the present dissertation, two lines of research will be discussed. Both conversational implicatures and conventional implicatures will be addressed. For both types of implicature, we were interested in the processing costs underlying the derivation of these inferences, but also in the effect of specific task characteristics. Both types of implicature are investigated from a developmental viewpoint. Since there is already abundant literature available on conversational implicatures, only one of the six studies discussed in this dissertation will concern scalar implicatures. The five studies on conventional implicatures are more explorative in nature. Before investigating the possible cognitive costs associated with conventional implicatures, we were interested on a more general level whether these inferences are understood and what factors influence the derivation of these implicatures.

Chapter 2

Chapter 2 of this doctoral dissertation describes an experimental study on the scalar inference from *some*. We focused on three factors that might influence competence with these implicatures.

First, we were interested in the effect of age. Previous studies (e.g., Chierchia et al., 2001; Papafragou & Tantalou, 2004) have already shown that young children are capable of deriving scalar implicatures under the right experimental circumstances. However, children as young as three years have never been investigated, so we included

this age group in our experiments. Apart from three-year-olds, also five-year-olds and seven-year-olds participated in our experiments.

Second, Pouscoulous, Noveck, Politzer and Bastide (2007) showed that the nature of the task affects the number of scalar implicatures derived in an implicature task. However, because they manipulated a whole range of variables, they weren't able to disentangle the effect of task from other possible contributing factors. In our study we directly compared an action task with a linguistic task and these tasks were made similar in design. We expected an action task to be more accessible for deriving scalar implicatures than a linguistic task, which is assumed to be more difficult. Moreover, we not only looked at differences in the kind of task but also directly compared two identical tasks with a different (cognitive) content. Two truth-value judgment tasks were presented of which one contained simple visual stimuli (marbles and boxes) whereas the other task appealed on knowledge of the world. We were interested to see whether this specific content would influence the number of pragmatic inferences.

Third, we were interested in working memory involvement in deriving these scalar implicatures from *some*. In previous studies, cognitive resources had been manipulated by burdening working memory with a secondary task. This reduced the available cognitive resources that were assumed to be involved in scalar implicature processing. However, in our experiments in Chapter 2 we didn't manipulate working memory load but we directly measured working memory capacity. The same kind of reasoning lies behind this: children with a lower capacity have less capacity available than children with a higher capacity so we expected less pragmatic answers in the low working memory span group than in the high span group.

Chapter 3

Chapter 3 describes this dissertation's first study in a whole range of experimental studies on the conventional implicature from *but*. This primary study was carried out with adults and tried to reveal whether the conventional implicature from *but*, *so* and *nevertheless* is indeed understood. Participants were presented with short stories ending with a 'p but q' sentence. *P* and *q* were always contrastive arguments for a certain decision or conclusion. Next, participants were presented with either two *nevertheless*-conclusions (i.e. '*nevertheless* conclusion from *p*' and '*nevertheless* conclusion from *q*') or with two *so*-conclusions (i.e. '*so* conclusion from *p*' and '*so*

conclusion from *q'*). They were instructed to indicate the appropriate conclusion from the two options. The appropriate conclusion following *so* is the conclusion from the *q*-argument whereas the appropriate conclusion introduced by *nevertheless* is the conclusion from *p*. In our 'p but q' sentences, the content of the *p*– and *q*-arguments was manipulated: both sensible and irrelevant arguments were presented. The aim of this manipulation was to investigate whether the conventional meaning of *but*, *so* and *nevertheless* is understood irrespective of the content of these arguments. In Experiment 2 of this study, we explicitly asked participants to motivate their answer. This enabled us to see whether the content of the arguments is mentioned whenever the participants didn't provide the appropriate answer.

Chapter 4

In Chapter 4, an experimental study on children's competence with conventional implicatures will be discussed. A group of 8-to-12-year-olds performed the same conventional implicature task as described in Chapter 3. Additionally, children's working memory capacity was measured in order to investigate whether a higher working memory is related to a better understanding of the conventional meaning of *but*, *so* and *nevertheless*.

Chapter 5

Chapter 5 describes a conventional implicature study with children aged 8 to 12 as well but with a different response format than the one used in Chapter 3 and Chapter 4. Inspired by Katsos and Smith (2010) and Katsos and Bishop (2011), a three-point scale was used instead of a binary judgment response format. A scale has the advantage that it can reveal more insight in which factors are considered important when processing 'p but q' sentences, combined with *so* and *nevertheless*. We expected sentences in which there was a conflict between the conclusion inferred from *but* and a conclusion based on the content of the arguments (e.g., a 'p but q' sentence in which *p* is a sensible argument and *q* is an irrelevant argument and the conclusion from *so* has to be judged) to be represented as middle answers on the scale. In contrast, sentences in which all factors direct the reader towards the same conclusion (e.g., 'irrelevant *p* but sensible *q*. So *q*') were expected to be answered with the highest value on the scale (i.e. good conclusion).

Chapter 6

In Chapter 6, the role of working memory in processing conventional implicatures is addressed again. In this study, the role of cognitive processes is assessed by burdening adults' working memory with a secondary task. The main goal was to reveal whether a higher load on working memory would be associated with a drop in inferring the appropriate conclusions from 'p but q' sentences. Moreover, we slightly adapted our experimental stimulus set. First, we directly compared 'p but q' sentences with sentences in which *p* and *q* were separated by a period. This enabled us to see if the expected answer from *q* would be elicited significantly more often in 'p but q' sentences than in 'p . q' sentences. Second, we manipulated the strength of the *p*- and *q*-arguments instead of using sensible and irrelevant arguments which were used in the studies described in Chapter 3, Chapter 4 and Chapter 5. These manipulations created a more ecologically valid experiment.

Chapter 7

Chapter 7 describes a study with adult participants who were presented with 'p but q' sentences as a direct concession instead of an indirect concession. Moreover, one of the arguments always expressed a feeling of understanding towards the main character in the story (e.g., '*I understand that after many attempts you lost the hope for reconciliation, but a good communication between the two of you is important for the entire company*'). This represents a relevant argument in daily life and might even have repercussions for consoling talks. As in Chapter 6, 'p but q' sentences were compared with 'p . q' sentences. We were interested in the possible differences between these two types of sentences with regards to the argument in which the expression of understanding is expressed (*p* or *q*). The participants were instructed to judge on a seven-point scale how (mis)understood the main character in the story would feel.

Chapter 8

Finally, Chapter 8 summarizes the major conclusions of the two lines of research presented in this dissertation and provides a general discussion of these results with regards to three major topics that were investigated in this dissertation for both types of implicature, namely task characteristics, development and working memory involvement.

References

Bott, L., & Noveck, I.A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*(3), 437-457.

Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. In A. H.–J. Do, L. Dominguez & A. Johansen (Eds.), *Proceedings of the 25th Boston University Conference on Language Development* (pp.157-168). Somerville, MA: Cascadilla Press.

Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In S. Bresnan, J. Choe & J. McDonough (Eds.), *Proceedings of the 15th Annual Meeting of the North Eastern Linguistic Society* (pp. 94–110). Amherst, MA: Graduate Linguistic Student Association.

Crain, S., & Thornton, R. (1998*). Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology, 54*(2), 128-133.

Grice, H.P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics, Vol. 3* (pp. 41-58). New-York: Academic Press.

Grice, H.P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes, 20*(5), 667-696.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M.C. (1997). *Algemene Nederlandse Spraakkunst*. Groningen/Leuven: Wolters Noordhoff.

Horn, L.R. (2004). Implicature. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 3-28). Oxford: Blackwell.

Huang, Y. (2006). *Pragmatics*. New York: Oxford University Press.

Jasinskaja, K. (2012). Correction by adversative and additive markers. *Lingua, 122*, 1899-1918.

Katsos, N., & Smith, N. (2010). Pragmatic tolerance and speaker-comprehender asymmetries. In K. Franich, K.M. Iserman, & L.L. Keil (Eds.), *Proceedings of the*

34th annual Boston Conference in Language Development (pp. 221-232). MA, USA: Cascadilla Press.

Katsos, N., & Bishop, D.V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67-81.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature.* Cambridge, MA: MIT Press.

Musolini, J., & Lidz, J. (2002). Preschool logic: truth and felicity in the acquisition of quantification. In B. Skarabela, S. Fish, & A.H.-J. Do (Eds.), *Proceedings of the 26th Boston University Conference on Language Development* (pp. 406-416). Somerville, MA: Cascadilla Press.

Noveck, I.A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition, 78*(2), 165-188.

Noveck, I.A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*(2), 203-210.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition, 86*(3), 253-282.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition, 12*(1), 71-82.

Pouscoulous, N., Noveck, I.A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition, 14*(4), 347-375.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.

Van Belle, W. (2003). *Zwijgen is niet altijd toestemmen. De rol van inferenties bij het interpreteren en argumenteren.* Leuven: Uitgeverij Acco.

Van Belle, W., & Devroy, G. (1992). *Tegenstellende en toegevende connectoren. Een argumentatieve beschrijving* (Preprint 143, voorlopige publicatie). Katholieke Universiteit Leuven, Faculteit Letteren, Departement Linguïstiek.

Chapter 2

‘Some’: effects of age, task, task content and working memory

Abstract

In three experiments, we investigated the effect of age, task, task content and working memory (WM) on scalar implicature processing. We found that three-year-olds still prefer the logical meaning of *some* (*some* being compatible with *all*), but five-year-olds and especially seven-year-olds are highly competent pragmatic reasoners. Additionally we found that not only the nature of the task but also the specific task content influences the amount of pragmatic answers: an Action-Based Task (ABT) leads to more pragmatic answers than a metalinguistic Truth-Value Judgment Task (TVJT) that, in turn, leads to more pragmatic answers than a different TVJT that included more cognitive content. Finally, we found no effect of WM in both five-year-olds and seven-year-olds. Children with a high WM capacity did not provide significantly more pragmatic answers than children with a low WM capacity.

***Keywords**:* scalar implicature; working memory; task; task content; age

Janssens, L., Fabry, I., & Schaeken W. (submitted). ‘Some’: effects of age, task, task content and working memory. *Manuscript submitted for publication in Psychologica Belgica*.

Introduction

People communicate with each other to express what they feel, think, want, etc. Although this seems to happen effortlessly and automatically, the communication process involves more than just decoding a message between a messenger and a receiver. Apart from the literal meaning of a sentence, a speaker can convey a lot more by communicating an implicit meaning that is not explicitly expressed. The first systematic attempt to explain how these inferences are derived, belongs to Paul Grice. He offered a comprehensive framework of the mechanics of inferential communication (Grice, 1975). According to Grice, communication is a cooperative enterprise between people, governed by certain relational expectations about how a conversational exchange should be conducted. These relational expectations are called 'maxims' and Grice proposed four of these maxims: the Maxim of Quantity, the Maxim of Quality, the Maxim of Relation and the Maxim of Manner. These maxims respectively imply that interlocutors are always expected to offer contributions which are informative, truthful, relevant to the goals of the conversation and appropriately phrased. Grice introduced the term *implicature*, which refers to the meaning that is implied by the speaker but not explicitly stated.

Considerable experimental research has been devoted to *scalar implicatures,* i.e. implicatures based on the existence of ordered terms on a scale of informativity (e.g., <*all, most, many, some*>, <*always, often, sometimes*>, etc.). A scalar expression such as *some* can be interpreted in two different ways: either pragmatically ('some but not all') or logically ('some and perhaps all'). Whenever a weaker term (e.g., the quantifier *some*) is used, the general consensus is that a stronger term from the same scale (e.g., *all*) does not hold because the speaker did not use the stronger term. If the stronger term was applicable then the speaker would have been underinformative.

Experimental investigations into children's interpretation of scalar terms have concluded that preschool children are often insensitive to scalar implicatures in tasks involving language comprehension (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Noveck, 2001). In these studies, children seemed to attend only to the logical/semantic meaning of the scalar terms, even though they were shown to be linguistically competent. For example, Noveck (2001) found that 89% of the seven-to-eight-year olds in his study agreed with statements such as '*Some* giraffes have long necks.' Noveck (2001) concluded that "younger, albeit competent reasoners, initially treat a relatively

weak term logically before becoming aware of its pragmatic potential", and that, in this respect, "children are more logical than adults" (Noveck, 2001, p. 165).

The availability of cognitive resources is often used to explain this typically found pragmatic delay in children. As suggested by Noveck (2001), a plausible explanation for this delay is that inferring scalar implicatures requires effort and that children have less cognitive resources available than adults. There are two major theories with opposite predictions regarding this issue. According to the default theories (e.g., neo-Gricean theories; e.g., Levinson, 2000), implicature production happens automatically and only its inhibition demands processing costs. Contextual theories (e.g., Relevance Theory; Sperber & Wilson, 1995) in contrast, suggest that an implicature will only be produced if it is relevant in the context and they state that this production requires additional processing costs. Evidence in favor of Relevance Theory, regarding scalar implicatures, has been presented among others by Noveck and Posada (2003). Their experiments indicated that pragmatic answers require more time than logical answers. Assuming that longer time is associated with more processing costs, this provides indirect evidence for Relevance Theory.

In contrast to research showing that children initially reason logically, there is also substantial experimental evidence that children are not incapable of drawing scalar inferences and that they are aware of the pragmatic potential of scalar expressions. In this kind of studies, the prime interest is to discover what conditions facilitate implicature production for children. A key factor seems to be the nature of the task. For instance, Foppolo, Guasti, and Chierchia (2004) conducted experiments concerning the quantitative scale <*all, some*> using two different tasks: a Truth-Value Judgement Task (TVJT) (Crain & Thornton, 1998), in which participants had to decide whether (underinformative) statements were true or false, and a Felicity Judgement Task (FJT) (Chierchia et al., 2001). In the FJT, participants were presented with a pair of utterances with the same truth-value but different levels of appropriateness and were asked to choose the most felicitous description. When five-year-olds completed the FJT, the number of pragmatic responses was 95% while the number of pragmatic responses on the TVJT was only 50%.

Pouscoulous, Noveck, Politzer, and Bastide (2007) also examined the role of the nature of the task. In their first experiment, they replicated earlier findings showing that nine-year-olds were more likely than adults to consider as true statements such as '*some*

turtles are in the boxes' (uttered when *all* turtles are in the boxes) in a TVJT. In their second experiment, they presented an Action-Based Task (ABT), in which participants did not have to give a metalinguistic evaluation of statements but had to respond by performing an action. Children were presented with five boxes and five tokens. They were asked to adapt the situation to make it compatible with a statement. For example, if they were told 'I would like *all* the boxes to contain a token' and two of the five boxes already contained a token, they were expected to put a token in every empty box. The results showed that, when children were asked to perform an action rather than give a metalinguistic truth evaluation, the number of derived implicatures in children increased.

In the present study we build on these experiments by Pouscoulous et al. (2007). In Experiment 1, we compare pragmatic processing between a group of three-year-old children and a group of five-year-old children. To our knowledge, no scalar implicature research has been done with children under the age of four.

Experiment 1

In our first experiment, both three-year-olds and five-year-olds were tested. Our primary goal was to test such young children's pragmatic competence by means of two different implicature tasks. It has been shown that the nature of the task has an influence on the number of pragmatic answers in children (e.g., Pouscoulous et al., 2007) so we included two different tasks based on Pouscoulous et al. (2007). We made two important changes to the Pouscoulous et al. (2007) study. First, we presented the same group of children with both a TVJT and an ABT: manipulating the nature of the task within subjects allows direct comparison between the two tasks. Second, there was an important difference in content between the ABT and TVJT used by Pouscoulous et al. (2007). Whereas the ABT in Pouscoulous et al. (2007) only used tokens and boxes, in the TVJT, the children were presented with three types of animals that remained in front of them throughout the task. For each statement, they had to focus on one type of animal and ignore the other animals. Since the statements were randomly ordered, they constantly had to switch their attention between the three types, which placed greater demands on information processing than in the ABT. To remedy this problem, we made the two tasks more similar in design by using the same scenarios with marbles and boxes in both tasks.

We had two hypotheses. First, we expected to find an age effect: we expected the five-year-olds to be more pragmatic on the critical items than the three-year-olds. Second, we expected to find an effect of the nature of the task. We expected the ABT to be easier and therefore to lead to more pragmatic answers than the TVJT.

Method

Participants

The sample comprised 20 three-year-olds (14 boys and six girls) between the ages of 36 and 52 months with a mean age of 44 months (*SD*=4.7) and 23 five-year-olds (12 boys and 11 girls) between the ages of 55 and 71 months with a mean age of 62 months (*SD*=5.04). They were recruited from a primary school in Belgium (Sint-Annaschool, Duisburg). All were native Dutch speakers, including some bilingual children.

Action-Based Task (ABT)

The ABT consisted of three scenarios, each involving five plastic boxes and five marbles. In the 'All-scenario', all five boxes contained a marble. In the 'None-scenario', all the boxes were empty. In the 'Subset scenario', two boxes contained a marble. In each scenario, a puppet, handled by the experimenter, was used to utter the same four requests: 'I would like *all* the boxes to contain a marble' (*'Ik zou willen dat er in alle dozen een knikker zit'*), 'I would like *some* boxes to contain a marble' (*'Ik zou willen dat er in sommige dozen een knikker zit'*), 'I would like *none* of the boxes to contain a marble' (*'Ik zou willen dat er in geen van de dozen een knikker zit'*) and 'I would like *some* boxes *not* to contain a marble' (*'Ik zou willen dat er in sommige dozen geen knikker zit'*). This amounted to a total of 12 requests. The participants were instructed to make changes to the scenario to comply with the puppet's requests. For example, if the puppet said 'I would like *all* the boxes to contain a marble' in the 'Subset-scenario', the child was expected to put a marble in the three empty boxes.

There were two critical situations and 10 control statements. The first critical statement occurred in the 'All-scenario' when the puppet stated 'I would like *some* boxes to contain a marble'. If the child interprets *some* logically, he or she will make no changes to the scenario. However, if the child grasps the implicature, he or she will take at least one and maximum four of the marbles away. The second critical statement occurred in the 'None-scenario' when the puppet uttered the statement 'I would like *some* boxes *not* to contain a marble'. In this case, if the child interprets the statement logically, no action

should be taken. A pragmatic interpretation on the other hand would require an action (adding at least one and maximum four marbles to the boxes).

For the 10 control statements, there was a distinction possible between pragmatic and logical interpretations, only for the *some (not)* sentences. For example: when the request 'I would like *some* boxes to contain a marble' is uttered in the 'None scenario', a wrong answer would be to change nothing, a pragmatic answer would be to add one to four marbles and a logical answer would be to put a marble in every box. All other control sentences were either right or wrong; E.g., 'I would like *all* the boxes to contain a marble' in the 'None-scenario'. In this case the child is expected to put a marble in all five empty boxes. All other actions would be wrong.

Truth-Value Judgment Task (TVJT)

The children were presented with five boxes and five marbles in the three same scenarios as in the ABT. In each scenario, a puppet made the same four statements (amounting to a total of 12 sentences): '*All* the marbles are in the boxes' (*'Alle knikkers zitten in de dozen'*), '*Some* marbles are in the boxes' (*'Sommige knikkers zitten in de dozen'*), '*None* of the marbles are in the boxes' (*'Geen van de knikkers zit in een doos'*) and '*Some* marbles are *not* in the boxes' (*'Sommige knikkers zitten niet in de dozen'*). After each statement, participants had to decide whether the statement was true or false. The two critical statements were '*Some* marbles are in the boxes' in the 'All-scenario' and '*Some* marbles are *not* in the boxes' in the 'None-scenario'. In both cases, 'true' would be the logical answer, whereas 'false' would be the pragmatic answer.

The other 10 statements were control statements (e.g., '*Some* marbles are in the boxes' in the 'Subset-scenario'). These statements could only be answered right or wrong, in contrast to the control statements of the ABT.

Procedure

Each participant was interviewed individually for about 20 minutes. For both age groups, the order of the two tasks was randomized, so that half of the participants started with the TVJT and the other half with the ABT. In both tasks, the experimenter used a puppet called Knorrie. In the TVJT, the children were informed that the puppet sometimes says things that are correct and sometimes says things that are wrong. In the ABT, the children were told that the puppet would give requests regarding the boxes and the marbles and that they would either have to remove marbles, add marbles, or make no changes. Before the start of the experiment, the children were given three

practice questions in the ABT. These questions were very similar to the experimental sentences but employed numbers instead of quantifiers. The three training questions were: 'I would like two boxes to contain a marble', when only one box contained a marble, 'I would like three boxes to contain a marble', when three boxes contained a marble and 'I would like two boxes to contain a marble', when three boxes contained a marble. These training questions were constructed so that the participants had to add marbles, change nothing and remove marbles. This way, they got acquainted with all types of actions they would have to perform during the experiment. If the children made errors on these training questions, the experimenter corrected them and explained their mistakes.

Results

Our first hypothesis concerned an effect of age. We expected five-year-olds to be more pragmatic than three-year-olds. Our second hypothesis concerned an effect of the nature of the task. We expected that the ABT would lead to more pragmatic answers than the TVJT.

With regards to our first hypothesis, we first compared the two age groups concerning the number of pragmatic answers they provided. However, in our analyses of the critical items of the ABT, five three-year-olds were excluded. These children didn't provide logical or pragmatic answers on the critical items of the ABT, but simply wrong answers (i.e. taking *all* marbles away when they were asked 'I would like *some* boxes to contain a marble' in the All-scenario and/or putting a marble in *all* the boxes when they were asked 'I would like *some* boxes *not* to contain a marble' in the None-scenario). These five children were included in all other analyses.

We found that the three-year-olds were significantly less pragmatic than the five-year-olds on both the ABT (Mann-Whitney U Test, n_1=15, n_2=23, U=98.5, Z=-2.43, p=.008) and the TVJT (Mann-Whitney U Test, n_1=20, n_2=23, U=119, Z=-2.94, p=.002). The three-year-olds provided 46.7% (ABT) and 45.0% (TVJT) pragmatic answers compared to 80.4% (ABT) and 76.1% (TVJT) for the five-year-olds. We also found that the older children provided more correct answers on the control items than the younger children on both tasks (ABT: 97.4% vs 88.5% correct answers; Mann-Whitney U Test, n_1=20, n_2=23, U=124.5, Z=-2.98, p=.002; TVJT: 93.9% vs 84.5% correct answers; Mann-Whitney U Test, n_1=20, n_2=23, U=111.5, Z=-3.07, p=.001).

With regards to our second hypothesis, we analyzed the two age groups combined. We found that the ABT was significantly easier than the TVJT since it led to more correct answers on the control sentences (93.3% vs 89.5% correct answers; Wilcoxon Signed Ranks test, n=26; T=88.5; p=.01). The easier ABT also led to more pragmatic answers (67.1%) than the TVJT (61.5%) but this difference was not significant (Wilcoxon Signed Ranks test, n=18; T=106.0; p=.17).

When we look at the difference between the two tasks for the two age groups separately, we only found a significant difference between the control sentences of the ABT and the TVJT for the five-year-olds (Wilcoxon Signed Ranks test, n=13; T=23; p=.049) and a marginally significant difference for the three-year-olds (Wilcoxon Signed Ranks test, n=13; T=23; p=.051). However, there were no significant differences between the two tasks regarding the underinformative sentences, for either of the two age groups separately.

The 10 control sentences of the ABT included four requests (the requests with *some (not)*) that could be answered in three different ways; either wrong (e.g., taking all marbles away when the request was 'I would like *some* boxes to contain a marble'), pragmatically (e.g., placing one to four marbles in the boxes when the request was *some*), or logical (e.g., placing a marble in *all* the boxes when *some* was requested). In our analyses above, we scored both the pragmatic and the logical answer as correct. However, when we look at the different types of answers separately, we find a significant difference between our two age groups. The three-year-olds provided 45.0% logical and 45.0% pragmatic answers on these sentences compared to 7.6% logical and 90.2% pragmatic answers for the five-year-olds (X^2=41.1, df=2, p<.001).

Discussion

Our results confirmed that there is an effect of age in pragmatic competence. Three-year-olds still prefer a logical interpretation of the scalar term *some* over a pragmatic one, whereas the majority of the five-year-olds favors a pragmatic interpretation. This difference was also clear from the four control sentences that could be answered logically in the ABT. About half the time, the three-year-olds spontaneously produced a logical answer, whereas the five-year-olds practically never did.

We made a distinction between two different tasks because we expected –in accordance with Pouscoulous et al. (2007)- the ABT to lead to more pragmatic answers

than the TVJT. We only found a reliable difference between the two tasks regarding the number of correct answers on the control sentences. As expected, the ABT was easier than the TVJT, but it did not lead to significantly more pragmatic answers.

Because we found evidence that pragmatic competence increases with age, this indirectly supports the assumption that pragmatic reasoning requires cognitive resources. As Pouscoulous et al. (2007) suggested, cognitive resources are important in implicature production and may explain why easier tasks, that require less cognitive resources, lead to more pragmatic answers than more difficult tasks. In adults, it has been shown that burdening WM decreases implicature production by 10% (De Neys & Schaeken, 2007). Moreover, Dieussaert, Verkerk, Gillard, and Schaeken (2011) found an interaction between cognitive load and WM capacity that influences pragmatic reasoning. They measured participants' WM capacity by means of the Operation Span Task for group testing and created three WM groups based on the performance on this WM task: the low-, middle- and high span group. They found an effect of cognitive load, only for the participants with a low WM capacity. The low span group provided fewer pragmatic answers when WM was burdened with a secondary task. The middle- and high span groups' pragmatic answering was not influenced by the cognitive load. This finding, that especially low capacity people are influenced by cognitive load, leads to the assumption that an effect of WM should be found in children's pragmatic reasoning because children's cognitive resources are limited. So far, no research has been conducted on children that directly investigated the role of cognitive resources.

Based on the findings of De Neys and Schaeken (2007) and Dieussaert et al. (2011), it can be assumed that people with less cognitive resources will be less pragmatic than people with more cognitive resources. In Experiment 2 we will measure WM capacity in five-year-old children and investigate whether children with a high WM capacity produce more scalar implicatures than children with a low WM capacity.

Experiment 2

Method

Participants

The sample comprised 48 five-year-olds (28 boys and 20 girls) between the ages of 62 and 73 months with a mean age of 67 months (*SD*=2.86), recruited from two different

schools in Belgium. None of these children participated in Experiment 1. All were native Dutch speakers.

TVJT, ABT

The same TVJT and ABT were used as in Experiment 1.

Working Memory Tasks

The children performed three WM tasks. First, the auditory (phonological loop) component was measured using the Digit Span Forward task in which subjects have to repeat an orally presented list of numbers. The list starts with a sequence of two numbers and keeps increasing until the child makes two errors within one block of the same digit length. Second, the visual component (visuo-spatial sketchpad) was measured using the Corsi Block Span test. In this test, the children were presented with nine wooden blocks on which the experimenter tapped a pattern and the children were instructed to repeat the sequence. The sequence becomes longer until the child makes two errors within one block of the same difficulty level. The third WM task, which was intended to provide a 'central executive' measure, was the Digit Span Backward task. This task is identical to the Digit Span Forward, except that the subject needs to repeat the sequence of numbers in reverse order. The raw scores for each of these tasks (i.e. the total number of correct answers) were converted into z-scores, which were then added up to compute the WM span.

Procedure

The procedure was exactly the same as in Experiment 1. The only difference was the extra WM measure. All children first completed the three WM tasks and next, the order of the other two tasks was randomized, so that half of the participants started with the TVJT and the other half with the ABT.

Results

Even though we did not find a significant difference between the ABT and the TVJT in Experiment 1, we hypothesized that there would be differences in implicature production and performance between the TVJT and the ABT. Our second hypothesis concerned an effect of WM.

Our first hypothesis about the difference in performance was confirmed by the finding that the TVJT leads to significantly more errors than the ABT on the control

statements (8.5% versus 1.5%, respectively; Wilcoxon Signed Ranks test, n=26; T=20.5; p<.001).

With regard to the critical sentences, we hypothesized that the ABT would lead to more pragmatic answers than the TVJT. Again, our hypothesis was confirmed. The children responded pragmatically to the critical sentences in 90.5% of the instances on the ABT, compared to 70.0% on the TVJT (Wilcoxon Signed Ranks test, n=20; T=22.5; p=.001).

Regarding our second hypothesis, we compared a high WM span group (N=16; M=2.13; SD=0.82) with a low WM span group (N=16; M=-2.37; SD=1.42) with regard to the number of correct answers on the control sentences and the number of pragmatic responses, for each of the two tasks. The results are displayed in Table 1.

While there were no significant differences in pragmatic processing, the number of correct responses to the unambiguous control sentences differed significantly between the two groups. The high span group was more accurate than the low span group on both the ABT (100% vs 95.6% correct answers; Mann-Whitney U test, n_1=16, n_2=16; U=96; Z=-2.099; p=.018) and the TVJT (93.8% vs 87.5% correct answers; Mann-Whitney U test, n_1=16, n_2=16; U=76; Z=-2.079; p=.019).

When we look at the number of logical answers on the control sentences of the ABT, we found that the low span group produced more logical answers than the high span group (10.9% vs 4.7%). However, this difference was not significant (X^2=1.74, df=1, p=.188).

Table 1 *Percentages of respectively correct and logical answers on the control sentences and critical sentences of the ABT and TVJT for low- and high WM span children (Experiment 2)*

	Control Sentences		Critical Sentences	
	Low span (n=16)	*High span (n=16)*	*Low span (n=16)*	*High span (n=16)*
TVJT	87.5	93.8	43.7	25.0
ABT	95.6	100.0	6.2	12.5

Discussion

In contrast to Experiment 1, the ABT did lead to significantly more pragmatic answers than the TVJT in Experiment 2. In addition, the five-year-olds made fewer mistakes on the ABT control statements than on the TVJT control statements. These results indicate that metalinguistic tasks are harder than tasks that don't require a verbal response. A possible reason why the difference in pragmatic reasoning was not found for the five-year-olds in Experiment 1 could be that the sample of children was too small.

The results of Experiment 2 show that five-year-old children are competent pragmatic reasoners. Their competence is still 'vulnerable', but taking into account certain factors such as task complexity, task content etc., they are capable of producing scalar implicatures on a high level. This confirms the findings of Pouscoulous et al. (2007). Moreover, the validity of our results was enhanced by manipulating the nature of the task within participants and by changing the design of the TVJT to make it more comparable to the ABT. This allows us to attribute the results to the task's cognitive demands and to conclude that the nature of the task is very important in implicature processing in five-year-olds.

Our WM measures revealed no significant differences in implicature processing between a group of low span children and a group of high span children. The high span children did make significantly fewer errors on the control statements of both tasks and were less logical on the control statements of the ABT (although this difference was not significant). Even so, these WM results do not allow us to draw firm conclusions about the role of WM in scalar implicature processing.

Remarkably, the five-year-olds in our experiments produced a much higher percentage of pragmatic answers than the children tested in Pouscoulous et al. (2007). They were equally pragmatic on the ABT and more pragmatic on the TVJT than the seven-year-olds and the adults in Pouscoulous et al. (2007), who conclude that "Only 7-year-olds reveal behavior that approaches that of adults among the standard cases and even among them adultlike implicature performance is less likely when it concerns negative sentences" (Pouscoulous et al., 2007, p.371).

Since the age of seven is mostly found to be the age at which children really begin to demonstrate pragmatic skills (e.g., Guasti et al., 2005), we ran the same experiment with a group of seven-year-olds. We expected them to be even more pragmatic than the five-year-olds. In addition to the ABT and TVJT used in Experiment 1 and Experiment 2,

we included a TVJT that is often used in experimental research on implicatures, i.e. the world-knowledge TVJT from Noveck (2001). By including this task, the children have to perform two different TVJT's that only differ in the specific content used. The TVJT that was also used in Experiment 1 and Experiment 2 involves simple materials (marbles and boxes) while the content of the other TVJT requires children to rely on their knowledge of the world. We expect this to be more difficult than the other TVJT.

Even though we did not find a significant WM effect in the five-year-olds of Experiment 2, we also measured WM in the seven-year-olds in Experiment 3. The WM tasks used in Experiment 2 were originally designed for children from the age of six (Working Memory Test Battery for Children (WMTB-C); Pickering & Gathercole, 2001). This means that the absence of a reliable WM effect might be attributed to the difficulty of the WM tasks that were used. These tasks should be suitable for seven-year-olds.

Experiment 3

Method

Participants

Thirty-four seven-year-olds (18 girls, 16 boys) between the ages of 6.9 and 8.5 with a mean age of 7.5 (*SD*=.32) participated in this experiment. All participants were recruited from the same school and were native Dutch speakers.

TVJT, ABT and WM Tasks

The same TVJT, ABT and three WM tasks were used as in Experiment 2.

World-knowledge TVJT

In order to investigate whether the specific content of the task plays a role in implicature production, the seven-year-olds conducted a task based on Noveck (2001; Experiment 3). In this task, the children were presented with 30 statements (translated into Dutch) and were instructed to indicate whether or not they agreed with each statement. The sentences were based on three types of information: factually universal, factually existential and absurd. The statements can be categorized in six subgroups:

(a) Five absurd *all* sentences (e.g., *all* birds have telephones.)
(b) Five absurd *some* sentences (e.g., *some* fish are made of leaves.)
(c) Five true *all* sentences (e.g., *all* elephants have trunks.)

(d) Five true (and felicitous) *some* sentences (e.g., *some* flowers are yellow.)

(e) Five false a*ll* sentences (e.g., a*ll* dogs have spots.)

(f) Five true (but pragmatically infelicitous) s*om*e sentences (e.g., *some* giraffes have long necks.)

We were particularly interested in the sentences from category (f). If children agree with such statements they are responding logically, while disagreeing implies a pragmatic response. If we look at the different types of statements, it is clear that switching quantifiers can make (c) interchangeable with (f) as well as (d) with (e). In this way, we created two versions of this task. In each version, both the *all* and the *some* sentences were randomized, as were the different types of statements.

Procedure

The procedure was exactly the same as in Experiment 2. However, an additional test was administered after all other tests were performed. All children received a paper with the 30 statements included in the world-knowledge TVJT. These statements were read out to them and they were asked to indicate, for each statement, whether they agreed or disagreed by circling the appropriate answer.

Results

We had two different hypotheses. The first hypothesis concerned an effect of the nature of the task. We expected the ABT to be easier than the TVJT that, in turn, we expected to be easier than the world-knowledge TVJT. Accordingly, we expected the ABT to lead to the most pragmatic answers and the world-knowledge TVJT to the least. Our second hypothesis concerned an effect of WM.

Regarding our first hypothesis, the TVJT control statements led to 95.9% correct answers, compared to 100.0% for the ABT (Wilcoxon Signed Ranks test, n=13, T=91.0, p<.001). For the control statements of the world-knowledge TVJT, the number of correct answers was 94.0% which differed significantly from the ABT (Wilcoxon Signed Ranks test, n=25, T=325.0, p<.001) and marginally significantly from the other TVJT (Wilcoxon Signed Ranks test, n=28, T=133, p=.055). Regarding the critical sentences, there were no significant differences between the TVJT and the ABT in the number of pragmatic answers (91.2% versus 94.1%, respectively; Wilcoxon Signed Ranks test, n=8, T=22.5, p=.24). In contrast, the world-knowledge TVJT only yielded 69.4% pragmatic answers,

which differed significantly from the other TVJT (Wilcoxon Signed Ranks test, n=23, T=229.5, p=.003) and from the ABT (Wilcoxon Signed Ranks test, n=22, T=34.5, p=.002).

Regarding our second hypothesis, we compared a group of high WM span children (N=11; M=2.32; SD=1.07) with a low span group (N=11; M=-2.38; SD=1.06). The results are displayed in Table 2. No significant differences were found between the two groups on any of the three tasks, neither in pragmatic responses, nor in performance on the unambiguous sentences.

Table 2 *Percentages of respectively correct and logical answers on the control sentences and critical sentences of the ABT, TVJT and world-knowledge TVJT for low- and high WM span children (Experiment 3)*

	Control Sentences		Critical Sentences	
	Low span (n=11)	*High span (n=11)*	*Low span (n=11)*	*High span (n=11)*
TVJT	96.4	95.5	0	13.6
ABT	100.0	100.0	4.5	0
World-knowledge TVJT	92.8	94.9	32.7	16.4

General Discussion

The three experiments reported in this article investigated pragmatic competence in young children. In Experiment 1, both three-year-olds and five-year-olds performed a metalinguistic TVJT and an ABT, in which children did not have to answer verbally. Children as young as three years had never been investigated in scalar implicature research. Our results showed that five-year-olds are competent pragmatic reasoners, interpreting *some* mostly pragmatically, whereas three-year-olds clearly favor the logical meaning of *some*. This indicates lack of pragmatic competence since their performance on the control sentences revealed overall linguistic competence with the quantors used in the tasks. The three-year-olds' preference for the logical meaning of *some* was also shown in the control sentences of the ABT that could be answered logically. The three-year-olds spontaneously provided the logical answer significantly much more than the five-year-olds.

Contrary to our expectations, Experiment 1 revealed no difference in the number of pragmatic answers between the two different tasks. Based on the findings of Pouscoulous et al. (2007), we expected the ABT to be easier than the TVJT and therefore to lead to more pragmatic answers. We did find a significant difference in the difficulty of the task (the control sentences of the ABT were answered more accurately than the control sentences of the TVJT) but the ABT did not lead to significantly more pragmatic answers than the TVJT.

In Experiment 2, a group of five-year-olds performed the same tasks as in Experiment 1. Additionally, a measure of WM was included. Based on the assumption that pragmatic reasoning requires cognitive effort, we expected an effect of WM. We expected children with a high WM capacity to be more pragmatic than children with a low WM capacity since they have more cognitive resources available. As in Experiment 1, we also wanted to test the hypothesis that the nature of the task plays an important role in implicature research. In contrast to Experiment 1, this hypothesis was confirmed in Experiment 2. We found, as expected, that a more difficult TVJT caused the children to be less accurate and less pragmatic than an ABT in which children did not have to answer verbally. This difference cannot be caused by a difference in task design because the two tasks were similar in design, but by a difference in task complexity. Manipulating the nature of the task is sufficient to show that, under the right circumstances, children as young as five years are capable of spontaneously producing implicatures. It is unclear why we did not find an effect of the nature of the task in Experiment 1. Since the three-year-olds showed very little pragmatic competence, we should look only to the group of five-year-olds. However, even if we only consider the group of five-year-olds in Experiment 1, no effect of the nature of the task can be found. It might be that the sample of five-year-olds was too small to find a significant effect.

We did not find any support for our hypothesis concerning WM. Five-year-olds with a high WM capacity were not significantly more pragmatic than those with a low WM capacity.

In Experiment 3, we investigated a group of seven-year-olds whom we expected to be even more pragmatic than the five-year-olds in Experiment 2. They performed the same tasks as in Experiment 2, including the WM tasks. Additionally, an extra task was administered: a TVJT based on world-knowledge that is often used in scalar implicature research (e.g., Noveck, 2001).

The expectation that seven-year-olds would provide even higher rates of pragmatic answers than five-year-olds was confirmed: the pragmatic response rate was so high that it did not lead to a significant difference between the ABT and the TVJT. However, when the children performed a TVJT involving world-knowledge statements, pragmatic responses dropped by 22%. For the world-knowledge TVJT, the children need to rely on the knowledge they have stored in their memory, whereas in the simple TVJT, they just have to rely on the boxes and marbles in front of them, which is less demanding on memory resources. Another difference between the two TVJT's that might influence pragmatic reasoning is that the TVJT with the marbles and the boxes is based on visual input (the marbles and the boxes) whereas the world-knowledge TVJT is not based on visual input.

This difference in the number of pragmatic answers between the two TVJT's shows that not only the nature of the task plays an important role in scalar implicature processing, but also the specific task content. The instructions of the two tasks were completely identical –indicating whether statements are wrong or right- but the content of the statements differed. The more cognitive world-based knowledge was required, the less pragmatic answers were provided. This cognitive content specifically affected pragmatic processing since the seven-year-olds proved to possess the world-knowledge required to judge the statements correctly by being highly accurate on the control statements of the world-knowledge TVJT.

The hypothesis that easier tasks lead to significantly more pragmatic answers than more difficult tasks is based on the assumption that cognitive resources are critical in implicature production (De Neys & Schaeken, 2007). As easier tasks require fewer cognitive resources than complex tasks, more cognitive resources remain available for producing implicatures. However, similar to Experiment 2, we did not find a reliable WM effect in the seven-year-olds. Even when we performed the WM analyses on the combined sample from Experiment 2 and Experiment 3 (with the highest scoring children in each experiment as the 'high group' and the lowest scoring children as the 'low group'), we did not find a significant WM effect. We did find that the high span children were more pragmatic than the low span children on the most difficult task in each experiment (the TVJT in Experiment 2 and the world-knowledge TVJT in Experiment 3). Although this trend can be observed in our WM data, we are unable to find a single significant WM effect. However, it is worth mentioning that even though we

would have expected a WM effect, the absence of a reliable effect is not that surprising given that the significant WM effect found in adults was only small (De Neys & Schaeken, 2007). This is necessary to ensure a smooth flow of communication.

A possibility for future research is to manipulate WM the same way De Neys and Schaeken (2007) did. Instead of measuring WM they presented participants with a secondary task in order to burden WM. A secondary task based on De Neys and Schaeken (2007), adapted for child use, might be a better method to investigate the role of WM in implicature production in children.

In sum, in three experiments we replicated the finding that there is a clear developmental trend in pragmatic competence. Three-year-olds show very little pragmatic competence whereas five-year-olds and especially seven-year-olds clearly understand and prefer the pragmatic meaning of *some*. A second important finding is that the nature of the task and the specific task content are very important in scalar implicature production in young children: more cognitive tasks or more cognitive task content cause a decrease in implicature production. It is important that this factor is taken into account when investigating implicature production in children because it can lead to wrongly drawn conclusions. Another factor that might need to be taken into account in future research is a measure of general language ability. Since it was found that metalinguistic tasks are harder than action tasks, it is plausible that general language ability may account at least partly for these results. It could be that, for such young children, general language ability is more important than WM capacity.

Finally, it is worth mentioning that pragmatic competence seems inextricably linked to people's mother language. Pouscoulous et al. (2007) had already shown that there is a difference in the number of pragmatic answers between the French *certains* and *quelques*, used as translations of *some*. Likewise, compared to other developmental implicature studies, it seems that Dutch speaking children are highly pragmatic when interpreting *some*.

References

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. In A. H.–J. Do, L. Dominguez & A. Johansen (Eds.), *Proceedings of the 25th Boston University Conference on Language Development* (pp.157-168). Somerville, MA: Cascadilla Press.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar. A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, MA: MIT Press.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology, 54*(2), 128-133.

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology, 64* (12), 2352-2367.

Foppolo, F., Guasti, M. T., & Chierchia, G. (2004). *Pragmatic inferences in children's comprehension of scalar items*. Talk presented at Second Lisbon Meeting on Language Acquisition, Lisbon, Portugal.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*, *Vol. 3* (pp.41-58). New York: Academic Press.

Guasti, M. T., Chierchia G. , Crain S., Foppolo F., Gualmini A., & Meroni L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes, 20*(5), 667–696.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition, 78*(2), 165-188.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*(2), 203-210.

Pickering, S. J., & Gathercole, S. E. (2001). Working Memory Test Battery for Children. Psychological Corporation UK.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs and implicature production. *Language Acquisition, 14*(4), 347-375.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.

Chapter 3

‘But’ how do we reason with it: An experimental investigation of the implicature stemming from ‘but’

Abstract

In two experiments, we aimed to investigate whether people truly make the inference induced by *but* – combined with *so* and *nevertheless* - in ‘p but q’ sentences constructed as distancing contrastive connections. In Experiment 1, our participants were presented with ‘p but q’ sentences that contained both sensible and absurd arguments and were instructed to indicate the appropriate *so*- or *nevertheless*-conclusion. We found that, while people do grasp the pragmatic meaning of *but*, the content of the arguments plays a very important role: when a sensible and an absurd argument were combined, the majority of participants based their answer on the sensible argument. We also found that the expected *nevertheless*-conclusions from *p* are very hard to make since they require participants to overrule the inference stemming from *but*. In Experiment 2 we further explored the role of the content by explicitly asking participants to explain/give reasons for their answers. Consistent with our expectations, we found, that whenever participants did not infer the appropriate conclusion from *but*, they referred to the content of the arguments. This means that people spontaneously consider the implication from the *p*-argument as well.

Keywords: conventional implicature; *but*; relevance theory; content

Janssens, L., & Schaeken, W. (2013). ‘But’ how do we reason with it: An experimental investigation of the implicature stemming from ‘but’. *Journal of Pragmatics, 57,* 194-209.

Introduction

The term 'implicature', which refers to what a speaker intends but does not state explicitly, was first introduced by Paul Grice, one of the founding fathers of pragmatics (Grice, 1989). Grice (1989) distinguishes two categories of implicatures, i.e. conversational implicatures and conventional implicatures. To date, experimental research on implicatures has almost exclusively focused on (generalized) conversational implicatures and more specifically on the subcategory of *scalar implicatures*. This type of implicatures is based on the existence of terms that are similar to each other in meaning, but differ in the amount of information they express (e.g., <*all, most, many, some*>, <*always, often, sometimes*>, <*certain, probable, possible*>, <*hot, warm*>, etc.). Upon hearing that '*some* of the books are red', most listeners will infer pragmatically that *not all* of the books are red. However, logically, the quantifier *some* is compatible with *all* and its literal, semantic interpretation is '*some and possibly all*'. Still, the general consensus is that, whenever a weaker term is used, a stronger term from the same scale does not apply because the speaker did not use a stronger term. The expectation that the speaker is as informative as possible elicits the pragmatic interpretation of the quantifier (e.g., '*some but not all'*).

In this study we will not focus on conversational implicatures, but on the far less frequently investigated conventional implicature. Drawing on a wide range of currently used definitions, Horn (2004) tries to develop an encompassing definition of this concept:

> "Unlike an entailment or logical presupposition, this type of inference is irrelevant to the truth conditions of the proposition. This inference is not cancellable without contradiction, but it is detachable, in the sense that the same truth-conditional content is expressible in a way that removes (detaches) the inference. Such detachable, but non-cancellable aspects of meaning that are neither part of, nor calculable from, 'what is said' are conventional implicatures." (Horn, 2004:4)

"Conventional implicatures are attached by convention to particular lexical items or linguistic constructions" (Huang, 2006:56). The following is an example of a conventional implicature:

(1) She is poor, but she is honest.

Conventionally, *but* implies a contrast (in this case between poor and honest). This is how people interpret *but*, even though this contrast is not explicitly expressed.

In our study, we investigated the implicature induced by the conjunction *but*. The 'p but q' utterances that we investigated were constructed as distancing contrastive connections. In the Algemene Nederlandse Spraakkunst (ANS; General Dutch Grammar), the distancing contrast is distinguished from the replacing- (*vervangende*) and the dividing (*verdelende*) contrast (Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn, 1997). In a dividing contrast, *but* can be replaced by *and*. When *but* is used instead of *and*, the contrastive connection is emphasized, but not vice versa (e.g., she is poor and she is honest). In a replacing contrast, the first clause is always a negation and the second clause seems to replace the first clause by formulating what is true (e.g., not yellow, but green is my favorite color). In a distancing contrast, *but* connects two parts of a complex speech act and the second part is disassociated from the first part, without denying what is being expressed in the first part (Van Belle & Devroy, 1992). In the 'p but q' construction, the speaker endorses or recognizes that *p* is true (Van Belle, 2003). Thus, the first part *p* is always characterized as a concession. However, the word *but* prevents the inference that would normally be derived from *p*. This can happen in two ways: The first possibility is that *q* contains a conclusion that contradicts the inference from *p* (*p (p → r), but q (q = not-r) (so not-r)*). For example (Van Belle, 2003): "The milk is sour (→ will not drink milk), but I drink it". The second possibility is that *q* consists of an argument that can be considered as stronger or more relevant than the inference from *p*. The inference from *q* overrules the inference from *p* (*p (p → r), but q (q → not-r) (so not-r)*). E.g.: "The milk is sour (→ will not drink milk), but I am very thirsty (→ will drink milk). (So I'll drink the milk)". These two subtypes of a distancing contrast are also distinguished terminologically. In the first construction a 'concluding *but*' introduces the contrastive connection, which is called a 'direct concession' (Van Belle & Devroy, 1992). The second construction includes an 'argumentative *but*' and is regarded as an 'indirect concession' (Van Belle & Devroy, 1992). The experiments in this paper only bear on the use of *but* as an indirect concession. Anscombre and Ducrot (1977) postulate three claims concerning this type of 'p but q' utterance:

1. *q* is always the argument with most weight and the 'p but q' construction must be viewed as a defense of not-*r*.
2. By uttering 'p but q', the speaker always expresses some kind of acceptance of *p*.
3. 'p but q' constructions are always aimed at cancelling a particular conclusion *r*.

In the example above, the *p*-argument is argumentatively oriented towards the conclusion *r* (will not drink the milk), whereas the *q*-argument is argumentatively oriented towards the opposite conclusion not-*r* (will drink the milk). Thus, the arguments *p* and *q* have the opposite argumentative orientation. The 'argumentative *but*' indicates that the second part should be considered as more relevant, but eventually it is up to the reader or listener to infer the opposite conclusion from the *q*-argument.

Concerning the notion of 'argumentative orientation', it needs to be pointed out that the argumentative orientation of an argument is determined by a positive or negative value that we ascribe to its content, called the 'axiological value' by Anscombre and Ducrot (1977). In the remainder of this article, we label arguments whose axiological value is oriented towards a positive conclusion as 'positive arguments' and their counterparts as 'negative arguments'. For example, one of the stories used in our experiments describes a boy in doubt about whether or not to go to the dentist. He says: "*I am afraid of the dentist, but I have a toothache*." In this example, the *p*-argument (afraid of the dentist) is the negative argument because it is oriented towards the negative conclusion (he will not go to the dentist), whereas the *q*-argument (toothache) is the positive argument because it is oriented towards the positive conclusion (he will go to the dentist).

The word *so*, following a 'p argumentative but q' utterance, introduces the expected conclusion from *q* (e.g., "The milk is sour, but I am thirsty. So I will drink it."). In contrast, Lepère (2008) argued in her master's thesis that, according to Van Belle (2003), the word *nevertheless*, used as a conjunctive adverb, following a 'p argumentative but q' utterance has the purpose to again reverse the argumentative orientation, thus directing the reader towards the conclusion implied by *p* (e.g., "The milk is sour, but I am thirsty. Nevertheless I will not drink it."). Note that in all these examples *nevertheless* is used to translate Dutch *toch* even though these two adverbs do not have the exact same meaning.

This suggested pragmatic meaning of *nevertheless* was also suggested for the German variant *doch* in Schmerse, Lieven and Tomasello (2014), who claimed that "the primary function of the accented variant of German *doch* is that of revising a common ground belief" (Schmerse et al., 2014:117). Likewise, "Fischer (2007) characterizes the core meaning of accented *doch* as adversative. According to her model, a hearer can construe the pragmatic pretext from an utterance with accented *doch* "through the negation of the situation described in the current utterance"" (Fischer, 2007:52 in Schmerse et al., 2014:119).

It is important to note that the argumentative strength of the words *but*, *so* and *nevertheless* should be completely unrelated to and basically prevail over the content of the arguments (Van Belle, 2003). When the arguments trade places, the opposite conclusion is reached, purely by the agency of these connectors.

As mentioned above, little empirical research has been done on conventional implicatures. The concept 'conventional implicature' was introduced by Grice to describe conventional aspects of meaning that are not truth-conditional. This can be clarified by looking at how Grice discusses *but*. We refer again to the example used above:

(2a) The milk is sour, but I am very thirsty.

(2b) The milk is sour and I am very thirsty.

But is not a truth-conditional connector. In truth-conditional semantics, (2a) has the same truth-conditions and consequently the same meaning as (2b). In other words, *but* would be equated with *and* in this context. However, *but* contributes to the meaning of (2a): it creates a contrast between the two parts of the utterance and it causes the conclusion implied by the first part to be denied by the second part of the utterance. It is this contribution of *but* that Grice describes as a conventional implicature. Although Grice did not elaborate on conventional implicatures in his conversational logic, this issue inspired the development of new theories.

The relevance theory developed by Dan Sperber and Deirdre Wilson (1995) builds on Grice's theory, but changes some essential aspects. According to Wilson and Sperber, human cognition is equipped for 'the maximization of relevance' (Wilson & Sperber, 1993, 2004). The relevance theory postulates that, other things being equal, the more cognitive effects an input has, the more relevant it is. However, the processing and

computation of these cognitive effects require mental effort. The theory claims that, for a hearer, the relevance of an utterance is inversely proportional to the effort required to derive its contextual effects. So, relevance is a positive function of the effects achieved and a negative function of the effort invested. According to relevance theory, communication holds the promise of 'optimal relevance': the listener can be confident that the interpretation which yields an adequate contextual effect at the lowest possible processing cost is in fact the correct interpretation.

Blakemore (1987) claims that "it would make sense for languages to have devices that are linked to the three ways of achieving relevance, that guide the hearer to the specific cognitive effects the speaker intends" (in Hall, 2004:220). In this line of reasoning and within the context of relevance theory, Blakemore (2002) developed a procedural analysis of *but* that can be summarized as follows:

> "To say that *but* means denial is to say that it encodes a constraint that triggers an inferential route involving *contradicting and eliminating* an assumption that is *manifest* in the *context*." (in Hall, 2004:220)

Thus, according to Blakemore, the connector *but* encodes a specific procedure. This procedural analysis contrasts with Grice's analysis of *but*. Within Blakemore's terminology and frame of reference, she postulates that Grice's view on *but* implies that this connector encodes a concept, in particular the concept 'contrast'.

Iten (2005) refined Blakemore's interpretation of *but*. She claims that the contradicted and eliminated assumption is only required to be accessible to the hearer. "What is meant by 'accessible' is that the individual is merely capable of entertaining the assumption, and not necessarily accepting it as true or probably true." (in Hall, 2004:224). According to Iten, the claim that this assumption needs to be manifest, is too strong. "An assumption is manifest to an individual at a given time if and only if he is capable of representing it mentally at that time and accepting it as true or probably true" (Sperber & Wilson, 1995:39). According to Iten (2005), the meaning of *but* is best described by saying that this connector encodes the following procedure: "what follows (q) contradicts and eliminates an assumption that is accessible in the context." However, this does not mean that the hearer necessarily needs to mentally represent this assumption before processing the clause introduced by the connector *but*. Once this clause has been processed, the hearer does have to grasp what kind of assumption it is

that the speaker expects him to 'eliminate'. If the hearer does not do this, he will consider the use of *but* as unacceptable.

Hall (2004, 2007) proposes an analysis of *but* in which this connector would encode the following procedure:

> "So, for the use of *but* to be acceptable, what follows has to be undermining a conclusion that the hearer could have drawn. By indicating that this potential inference is getting cut off by what follows, *but* may save the hearer effort in reaching the intended interpretation of the clause it introduces. The context provides some evidence compatible with a certain inference; an aspect of the interpretation of the *but*-segment contradicts the result of this inference, had it gone through, so *but* signals that the speaker doesn't want the hearer to draw some potential conclusion. My suggestion for the constraint encoded by *but* is then that *but* indicates that the hearer is to suspend an inference that would result in a contradiction with what follows, so diverts him from a conclusion that he could potentially have drawn." (Hall, 2004: 228)
>
> …
>
> "It's an account on which *but* targets the *inferential process* itself, rather than any identifiable *conclusion* of this inference." (Hall, 2004: 226)

With regard to *but* constructions as indirect concessions, Hall postulates that the clause introduced by *but* does not seem to eliminate an assumption, but merely seems to introduce an argument that points in a different direction. Hall (2004, 2007) identified a number of counterexamples which demonstrated that both Iten's (2005) and Blakemore's (1987) view are inadequate. For example:

(3) A: Do you think that we can trust him?
B: He is honest, but he is a Republican, so I don't know.

In B's answer, the first clause implies that 'we can trust him', whereas the second clause implies that 'we cannot trust him'. According to Hall (2004), if the assumption 'we can trust him' was eliminated when the hearer reaches the end of the second clause, hearing 'so I don't know' would cause the whole utterance to be processed again and B's answer would sound marked. However, Hall claims that there is nothing about the utterance in

the example that sounds marked, which indicates that the cognitive effect does not imply 'contradiction and elimination'. That is why Hall considers her analysis of *but* superior to the 'contradiction and elimination' view, at least concerning indirect concessions. "The implication of the second clause (we cannot trust him) does not entirely seem to replace the implication of the first clause (we can trust him). It just has more weight, and this is all that follows from the constraint I'm proposing." (Hall, 2004: 229).

In our experiments, we wanted to investigate whether people truly make the inference induced by *but*, *so* and *nevertheless*. Participants were presented with 'p but q' constructions and asked to indicate the appropriate conclusion introduced by either *so* ('p but q, so *r*' vs. 'p but q, so not-*r*') or *nevertheless* ('p but q, nevertheless *r*' vs. 'p but q, nevertheless not-*r*'). The *p*- and *q*-arguments used were either sensible or absurd. In the example described above about the boy deciding whether or not to go to the dentist, both the toothache and the boy's fear of the dentist are sensible arguments (for going or not going to the dentist, respectively). In contrast, 'I am wearing a red sweater' would be an absurd argument in this context, as it is unrelated to the topic. Such absurd arguments were included in the experiments in order to determine whether people are influenced by the content of the arguments rather than by the structure of the sentence.

If it is true that the implication from *p* (*r*) is cancelled by using *but*, then we should expect few mistakes on the *so*-conclusions from 'p but q' constructions. If the cognitive effect truly involves 'contradiction and elimination', then the instruction conveyed by *but* (the *so*-conclusion from *q* follows, regardless of the content of the arguments) should be easy to follow. It would not be possible to derive an inappropriate *so*-conclusion from *p*. In contrast, according to Hall's (2004, 2007) procedural analysis of *but*, a *so*-conclusion from *p* does seem plausible in those cases where the content of the arguments gives rise to this. She postulates that, in the constructions relevant to our experiments, the implication from *p* is not entirely replaced by the implication from *q*; *q* merely has more weight.[1] By using the word *but*, the speaker suggests that *q* has more weight than *p*, but the *so*-conclusion from *q* does not necessarily need to follow in all cases. The content of the arguments and the context also contribute to the argumentation, can induce doubt (the utterance "so I don't know" can follow 'unmarked'

[1] In direct concessions, where *q*=not-*r*, the implication from *p* is entirely replaced by the implication from *q*. However, these instances are not relevant here because we only worked with indirect concessions.

from a 'p but q' construction) and can even play a decisive role. That is why *so*-conclusions from *p* seem to be plausible in some cases, which implies that people do not need to consider such conclusions, which do not-conform to the word *but*, as incorrect.

Concerning the *nevertheless*-conclusions, none of the authors mentioned above, nor any other author within the field of relevance theory has analyzed this word in this context. The experiments discussed in this paper can shed light on the conventional meaning of *nevertheless*. Based on Lepère (2008) and Van Belle (2003) (and also see Fischer, 2007; Schmerse et al., 2014), we expect that *nevertheless* directs the reader/hearer towards the conclusion inferred from the *p*-argument in a 'p but q' construction. In fact, we expect that the conventional meaning of *nevertheless* entails that the inferred conclusion from *q* should be denied. This would require the reader/hearer to make two inferences: first, the inference from *but* needs to be made, followed by the inference from *nevertheless* that implies that the opposite conclusion (i.e. the conclusion from *p*) is the conclusion that needs to be followed in this construction. From the 'contradiction and elimination' perspective, this would mean that the hearer needs to deny the implication from *p* (*r*) and replace it with the implication from *q* (not-*r*), followed by the implication from *nevertheless* that this implication *r* needs to become accessible again and accepted as the only correct conclusion. Consequently, from this view, *nevertheless*-conclusions from *p* should require more effort to make than *so*-conclusions. In Hall's (2004, 2007) analysis, it seems less effortful to derive a *nevertheless*-conclusion from *p* in a 'p but q' construction, since the hearer has not fully replaced the implication from *p* with the implication from *q*. According to Hall, by using *but*, the speaker merely indicates that the argument in the second clause has more weight.

Since we also included absurd *p*- and *q*-arguments we would expect, in line with Hall (2004, 2007), an interaction between the type of argument (absurd or sensible) and the type of conclusion (*so* or *nevertheless*): *nevertheless*-conclusions should be easier when the *q*-argument is an absurd argument, i.e. in the 'sensible but absurd' cases, whereas *so*-conclusions would be more difficult in these cases. On the basis of Hall's analysis, we would expect people to pay more attention to the content of the arguments whenever an absurd argument is included. This would lead them to base their conclusion on the sensible *p*-argument in the 'sensible but absurd' cases (i.e. the appropriate *nevertheless*-conclusion). For the same reason, we would expect these cases

to lead to more inappropriate *so*-conclusions because the appropriate *so*-conclusion is based on the *q*-argument.

Experiment 1

Method

Participants

In our first experiment, 63 undergraduate students from the University of Leuven (Belgium) participated (mean age of 19.6) in exchange for course credit. They were all native Dutch speakers.

Materials

The participants were presented with 18 context stories –all in Dutch- that described a person in doubt. For example:

> *Peter's friends decide to have a hamburger for lunch. Peter is unsure whether or not to join them.*

After this introduction, the person in doubt gives two contrastive arguments separated by *but*. For example:

> *Peter says: "I am hungry, but I already have other plans."*

The arguments included could be either absurd or sensible. In the example above, both arguments are sensible. The first, positive argument is oriented towards a positive conclusion ('Peter will join his friends'). The contrastive argument, introduced by *but*, suggests a negative outcome ('Peter will not join his friends').

These arguments were collected and rated through a number of pilot studies. In a first pilot study, nine students generated as many reasons as they could think of why the leading characters in the stories might make a certain decision or the opposite decision. In a second pilot study, 20 other students rated seven of the most frequently generated arguments for each story on a seven-point scale ranging from 1 (worst argument) to 7 (best argument) This allowed us to collect the most sensible arguments.

In a third pilot study, we generated absurd arguments, unrelated to the topic of the stories. For instance, in the example with Peter:

Peter says: "I am hungry, but I have curly hair."

A group of 19 participants rated the soundness of these arguments on a five-point scale ranging from 1 ('very bad argument') to 5 ('very good argument'). The highest mean rating was 2.3, while all other absurd arguments had a mean rating below 2. This means that all items were appropriate absurd arguments that could be used in the experiment.

The absurd arguments do not have a positive or negative axiological value, i.e. they are not oriented towards a particular positive or negative conclusion. In total, there were six combinations of arguments: 'positive sensible but negative sensible', 'negative sensible but positive sensible', 'absurd but positive sensible', 'absurd but negative sensible', 'positive sensible but absurd' and 'negative sensible but absurd'. There was no 'absurd but absurd' combination because absurd arguments do not have an axiological value. If both arguments are unrelated to the topic, there is no way of knowing what the conclusion should be.

After participants had read the 'p but q' construction, they were asked to choose the appropriate conclusion. They were told to indicate the conclusion that the person in the story would make, based on the construction of his utterance. They were explicitly told not to take into account the conclusion they themselves would make or the one that makes most sense. For half the stories they had to indicate the appropriate *so*-conclusion ('so *r*' or 'so not-*r*') and for the other half of the stories they had to indicate the appropriate *nevertheless*-conclusion ('nevertheless *r*' or 'nevertheless not-*r*'). For instance:

1. *"So, I will join my friends for lunch."*
2. *"So, I will not join my friends for lunch."*

or

1. *"Nevertheless, I will join my friends for lunch."*
2. *"Nevertheless, I will not join my friends for lunch."*

If *so* is used, the appropriate pragmatic conclusion is the conclusion following from the *q*-argument, whereas in the case of *nevertheless*, we expect that the appropriate pragmatic conclusion would be derived from the *p*-argument. For a full list of materials, see Appendix A.

Procedure

The participants were given a pen-and-paper task with the 18 stories and questions. There were 12 different versions of the task (six argument combinations x two conclusion options). Each participant was presented with three items from each argument combination and an equal number of *so*-questions and *nevertheless*-questions.

Results

In analyzing whether people make the inference induced by *but*, *so* and *nevertheless*, we first controlled if we should keep the distinction between positive and negative arguments in our further analyses. To that end, we compared the number of appropriate answers between the different categories. We found no significant difference in the number of appropriate answers when we compared 'positive p but negative q' with 'negative p but positive q' (68% and 63% appropriate answers respectively; X^2=.59; *df*=1; *p*=.44). Likewise, there was no significant difference between 'positive p but absurd q' and 'negative p but absurd q' (57% and 53% appropriate answers; X^2=.28; *df*=1; *p*=.60). Finally, the difference between 'absurd p but positive q' and 'absurd p but negative q' was also not significant (63% and 67% appropriate answers; X^2=.52; *df*=1; *p*=.47). As a result, we were allowed to collapse these categories in our further analyses. This produces the following combinations: 'sensible but sensible', 'absurd but sensible' and 'sensible but absurd'. The results from our analyses are displayed in Figure 1. We found that – when presented with two sensible arguments - participants gave the appropriate *so*-conclusion in 82% of the instances, compared to 48% for the *nevertheless*-conclusion (Wilcoxon Signed Ranks test, *n*=63; *T*=437.5; *p*<.001). Only the 'sensible but absurd' contexts yielded a high percentage of appropriate *nevertheless*-conclusions (78%). This differed significantly from 'sensible but sensible, nevertheless' (48%; Wilcoxon Signed Ranks test, *n*=63; *T*=526.5; *p*<.001) and from 'absurd but sensible, nevertheless' (31%; Wilcoxon Signed Ranks test, *n*=63; *T*=694; *p*<.001). On the other hand, the percentage of appropriate *so*-conclusions was only low in the 'sensible but absurd' contexts (37%). This differed significantly from 'sensible but sensible, so' (82%; Wilcoxon Signed Ranks test, *n*=63; *T*=46.5; *p*<.001) and from 'absurd but sensible, so' (97%; Wilcoxon Signed Ranks test, *n*=63; *T*=4; *p*<.001).

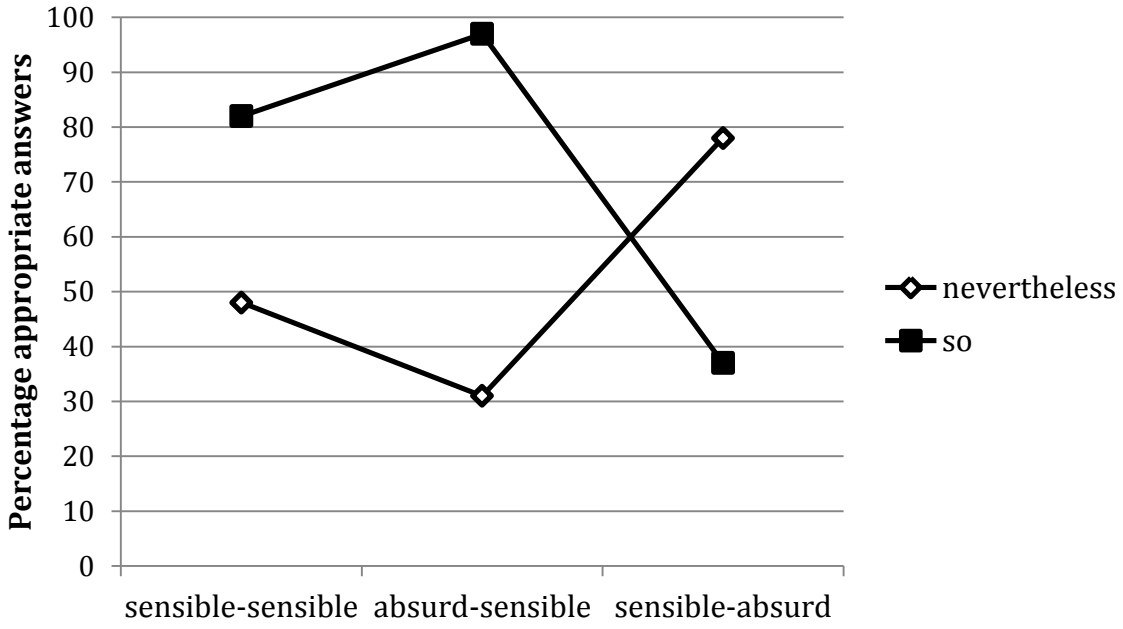


Figure 1. Percentages of appropriate answers for each of the argument combinations.

Discussion

The results show that, in general, people do make the inference induced by *but*. This can be seen in the high percentage of appropriate *so*-conclusions when two sensible arguments were presented. On the other hand, we found that the expected inference induced by *nevertheless* was not an obvious answer. When two sensible arguments were presented, only 48% expected *nevertheless*-conclusions from the *p*-argument were given. This is not consistent with our expectations. However, since the expected inference stemming from *nevertheless* contradicts the inference stemming from *but*, it is understandable that this inference requires more effort than the *so*-conclusion and therefore elicits a higher number of inappropriate answers.

That people sometimes do not make the inference implied by *but* is illustrated by the fact that, even when both arguments made sense, still 18% inappropriate *so*-conclusions were given. Most likely, in these situations, people pay more attention to the content of the arguments and base their conclusion on the most plausible argument rather than on the structure of the sentence (i.e. making the inference). The importance of the content was most evident in contexts in which absurd arguments were used. In most of these cases, participants chose the conclusion stemming from the sensible

argument regardless of the direction suggested by *but*, *so* or *nevertheless*. This would explain the high percentage of appropriate *nevertheless*-conclusions and the very low percentage of appropriate *so*-conclusions in the 'sensible but absurd' contexts.

If we try to relate these findings to the theories discussed in the introduction, the results concerning the *so*-conclusions seem to provide support for Hall's procedural analysis (2004, 2007). As mentioned above, Hall suggested that the implication from *q* does not entirely replace the implication from *p* but just has more weight. Consequently, a *so*-conclusion from *p* becomes more plausible if the content of the arguments gives rise to this. This is clearly the case in the 'sensible but absurd, so' contexts, in which participants preferred the inappropriate *so*-conclusions (from *p*). This finding and the finding that still 18% inappropriate *so*-conclusions are given in the 'sensible but sensible' contexts cannot be explained by the 'contradiction and elimination' view that argues that the implication from *p* is entirely replaced by the implication from *q*.

The results obtained for *nevertheless* are harder to interpret. As mentioned above, no one has analyzed the use of *nevertheless* in this context before. If *nevertheless* indeed entails that the inferred conclusion from *q* should be denied, then the expected appropriate inference from *nevertheless* would be the conclusion from *p*. However, we found that *nevertheless*-conclusions from *p* are very hard to make, even when two sensible arguments are used and the content of the arguments does not play a role. Thus, at first sight these results seem to contradict our expectation. However, we can speculate to what extent the theories discussed in the introduction can explain these results. For instance, these results seem plausible from a 'contradiction and elimination' perspective. In this view, the construction with *but* requires the hearer to deny the implication from *p* (*r*), remove it and replace it by the implication from *q* (not-*r*). Subsequently, this implication *r* needs to be made accessible again and accepted as the only right conclusion. Thus, from the 'contradiction and elimination' perspective, it seems plausible that this whole process requires mental effort and therefore leads to inappropriate answers. However, if it is that difficult to cancel the cancelled implication from *p*, then one should expect even lower percentages than the ones observed, i.e. percentages closer to 0%. Since the observed percentages are higher than can be expected from the 'contradiction and elimination' view, perhaps these findings cannot be explained by this account. Even more problematic is the fact that 78% appropriate *nevertheless*-conclusions from *p* were given in the 'sensible but absurd' contexts. This

cannot be accounted for by the 'contradiction and elimination' view. However, this finding might be accounted for by Hall's (2004, 2007) theory since this is a case in which the conclusion from *p* becomes more plausible when the arguments give rise to this. This being said, we stress again that our interpretations of *nevertheless* in the light of existing theories are only speculative since none of these theories made any predictions concerning *nevertheless*.

Experiment 1 showed that the content of the arguments seems to have a great effect on people's interpretations of *but*. This was further investigated in Experiment 2. If *nevertheless* indeed causes the reader to consider the implication from *p*, then merely the presence of *nevertheless* forced participants to consider this implication in Experiment 1. It is unclear how often this would happen spontaneously. Furthermore, if participants give these answers, are they indeed content-driven? Of special interest are the unexpected answers, based on the presence of *but*: do some participants spontaneously favor the *p*-implication in the 'sensible but absurd' constructions? If so, do these participants indicate a content-driven process when asked to justify their answer?

In Experiment 2, participants basically performed the same task as in Experiment 1 but instead of being instructed to give the appropriate *so*- or *nevertheless*-conclusion, they were simply asked to indicate the appropriate conclusion to be drawn from the *but* construction and to justify their answer. We hypothesized that people would refer to the content of the arguments whenever they selected the inappropriate answer, but mention the structure of the sentence when choosing the appropriate answer, especially in the difficult 'sensible but absurd' cases.

Experiment 2

Method

Participants

In our second experiment, 71 undergraduate students from the University of Leuven (Belgium) participated (mean age of 19.8) in exchange for course credit .They were all native Dutch speakers. None of them had participated in Experiment 1.

Materials

The same 18 contexts followed by a *but* construction were used as in Experiment 1. Again, both absurd and sensible arguments were used in the same six combinations. However, after each *but* construction, participants did not have to give the appropriate *so*- or *nevertheless*-conclusion but simply had to tick a 'yes' or 'no' box to the question 'Is the person in the story going to do whatever he (or she) is indecisive about?'. In addition, they were asked to justify their answer. See Appendix B for a full list of materials.

A (translated) example:

> *During the holidays, Sarah is staying with her grandmother for a couple of days. Sarah is sitting outside in the garden when her grandmother brings her a glass of lemonade. Sarah is unsure whether or not to drink the lemonade.*
>
> *Sarah says: "I don't like lemonade, but I am thirsty."*
>
> ➔ *Will Sarah drink from the glass of lemonade?*
>
> - *Yes*
> - *No*
>
> ➔ *Why do you think that?.....*

Procedure

The participants were given a pen-and-paper task with the 18 stories and questions. The various possible combinations of the arguments yielded six different task versions. Each participant was presented with three items from each argument combination.

Results

In the 'absurd but sensible' contexts, the percentage of appropriate answers was 93%, which differed significantly from the 'sensible but sensible' contexts (75%; Wilcoxon Signed Ranks test, n=71; T=36; p<.001) and the 'sensible but absurd' contexts (34%; Wilcoxon Signed Ranks test, n=71; T=16.5; p<.001).

To analyze participants' reasons for their answers, two independent raters were asked to label each specific justification as 'related to content', 'related to the structure of the sentence' or '?' (unrelated to either the content or the structure of the sentence). Interrater reliability was 0.92. On items where raters did not agree, they reached a

compromise so that each justification could be assigned to either the 'content' or the 'structure' category.

We hypothesized that people would refer to the content of the arguments whenever they selected the inappropriate answer, but mention the structure of the sentence when choosing the appropriate answer, especially in the difficult 'sensible but absurd' cases.

In the 'sensible but absurd' contexts, we were most interested in the reasons provided by participants who did make the correct inference (only 34%). Since the *but* construction led them towards the conclusion derived from the second argument, they had to ignore the first, sensible argument. We found that 84% of the justifications referred to the structure of the sentence. Some participants even tried to make sense of the absurd argument; for instance, in one story the argument that a teacher has a wife and children was given as a reason why this person should not win the 'teacher of the year' award in the construction "Mr. Van Damme is a good teacher, but he is married and has children." One participant answered that he would not win the award 'because he does not have much time to teach because of his wife and children'. This shows that the participant was aware that the answer should be 'no' because of the *but* construction and therefore tried to make sense of the absurd argument.

Eleven percent of the justifications consisted of newly invented arguments. Again, those participants seem to understand that *but* guided them towards a certain conclusion, so they tried to come up with a sensible argument that would be consistent with that conclusion. Finally, 5% of the justifications were inconsistent with the given answer.

In the 'sensible but sensible' contexts, the justifications given by participants who did make the expected inference were of less interest since both the content and the structure of the sentence might have led them to make this inference. This means that even if participants justified their answer by referring to the content of the arguments, this does not exclude that they were aware of the appropriate meaning of *but*.

However, we were interested in the justifications given by participants who did not make the expected inference, i.e. who drew a conclusion based on the *p*-argument in a 'p but q' construction. Among the justifications provided by participants who did not make the inference, 82% referred to the content of the arguments. Clearly, those participants preferred the *p*-argument over the *q*-argument. The other 18% of responses

consisted of justifications that neither referred to the content nor to the structure of the sentence. In these justifications, participants invented their own arguments or gave a reason that was inconsistent with their yes-or-no-answer. For example, in one story two boys ask their father to tell them a horror story before they go to bed. Their father thinks: 'The boys have been good, but they could get nightmares from a horror story.' When participants are asked "Will the father tell his sons a horror story?", one participant answered "Yes, because the boys asked it themselves". This is an example of a 'newly invented' argument.

Finally, in the 'absurd but sensible' contexts, both the content and the structure of the sentence guided the participants towards the inference, even more than in the 'sensible but sensible' contexts, because the sensible argument does not have to be weighed against another sensible argument but against an absurd argument and thus clearly has the upper hand. Hence, the reasons provided by participants who chose the appropriate response in these contexts were not of particular interest. Rather, we were interested in the justifications given by the small percentage of participants who did *not* make the correct inference. Among 25 such justifications, 20% were inconsistent (i.e. according to their justification they should have given the opposite answer), 48% were new arguments thought up by the participants and 32% consisted of negations of the *q*-argument (e.g., when the sensible *q*-argument was that something is expensive, one participant argued that it did not cost 'too much'). All these justifications could be said to be related to the content of the clauses.

General discussion

The goal of this study was to investigate whether people make the inference induced by *but*, *so* and *nevertheless* in 'p but q' sentences constructed as distancing contrastive connections. In our first experiment, we found that, in general, participants were aware of the appropriate pragmatic meaning of *but*. However, the content of the *p*- and *q*-arguments played a non-negligible role. If one of the arguments given is completely absurd, the majority of people will ignore the direction indicated by *but* and base their answer on the sensible argument. We also found that only 48% expected inferences from *nevertheless* were made when two sensible arguments were presented. This could indicate that the assumption that *nevertheless* directs the reader towards the implication *r* from *p* might not be correct. However, given the fact that, in order to make the

expected pragmatic inference with *nevertheless*, the inference from *but* needs to be overruled, it could be expected that *nevertheless*-conclusions require some effort to make and therefore lead to more inappropriate answers than *so*-conclusions. Furthermore, when we look at the results of Experiment 2, we found that 34% inferences from *but* were made in the 'sensible but absurd' contexts. This means that 66% of the answers were guided by the sensible *p*-argument. In Experiment 1, in which participants were asked to indicate the appropriate conclusion with *nevertheless*, 78% of the answers in the 'sensible but absurd' contexts were based on the sensible *p*-argument. Since this percentage is higher than in Experiment 2, in which *nevertheless* was not present, this seems to suggest that *nevertheless* does seem to have some additional influence on the interpretation of the sentence in the direction we expected.

In Experiment 2, we explicitly asked participants to give reasons for their answer. We did not ask for the appropriate *nevertheless*- or *so*-conclusion but simply for the appropriate conclusion stemming from *but*, which, in theory should always be the conclusion based on the *q*-argument. Again, we found evidence that the content of the arguments plays a very important role. For example, when we look at the difference between the 'sensible but sensible' and the 'absurd but sensible' condition, we found that the percentage of appropriate conclusions was 18% higher in the latter condition. Even though in both conditions the appropriate conclusion is derived from a sensible *q*-argument, it seems to matter whether this *q*-argument needs to be weighed against an absurd or a sensible *p*-argument: if weighed against an absurd argument, the *q*-argument clearly has greater weight, but when the *p*-argument is a sensible argument, this is less obvious. This indicates that people do spontaneously consider the implication from *p* and that this implication is not completely cancelled by the use of *but*. We found, as expected, that the consideration of this *p*-argument is content-driven: participants who reached the inappropriate *p*-conclusion tended to refer to the content of that argument or even to the oddity of the absurd argument (in the 'sensible but absurd' contexts) to justify their answer.

In sum, the results from these two experiments suggest that, when hearing a 'p but q' sentence, people generally do understand the pragmatic meaning of *but*, which results in a conclusion drawn from the *q*-argument. However, the content of the *p*- and *q*-arguments has an influence on the interpretation as well. Therefore, these results can be explained by Hall's theory (2004, 2007). She postulates that the clause introduced by

but does not seem to eliminate an assumption, but merely seems to introduce an argument that points in a different direction. The *q*-argument has more weight and is preferred over the *p*-argument but when the content of the *p*-argument allows it, a conclusion can be drawn from *p*. Indirectly, she thus stresses the importance of the content of the arguments, which is supported by our findings.

Our results with *nevertheless* do not offer any definite conclusions as to whether the assumption that *nevertheless* elicits the pragmatic inference from the *p*-argument holds or not. It could be that this assumption is true but that the low percentage of appropriate *nevertheless*-conclusions can be attributed to the fact that it is effortful to make this inference. In future research, it might be interesting to investigate how effortful it is to derive the inference from *but* and *nevertheless*. It has been found by many researchers that processing scalar implicatures is effortful and requires working memory (e.g., De Neys & Schaeken, 2007). From a developmental perspective, Noveck (2001) found that children answered less pragmatically than adults on a scalar implicature task, which was attributed to their cognitive capacities not being as developed as adults'. This is indirect evidence that deriving scalar implicatures requires mental effort. Given this evidence for conversational implicatures, one might wonder whether the same is true for conventional implicatures. This question will be addressed in future experiments by using our *but* task (1) in children, and (2) with working memory load.

References

Anscombre, Jean-Claude, Ducrot, Oswald, 1977. Deux 'mais' en français? Lingua 43 (1), 23–40.

Blakemore, Diane, 1987. Semantic Constraints on Relevance. Blackwell, Oxford.

Blakemore, Diane, 2002. Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers. Cambridge University Press, Cambridge.

De Neys, Wim, Schaeken, Walter, 2007. When people are more logical under cognitive load: Dual task impact on scalar implicature. Experimental Psychology 54 (2), 128-133.

Fischer, Kerstin, 2007. Grounding and common ground: Modal particles and their translation equivalents. In: Anita Fetzer & Kerstin Fischer (Eds.), Lexical Markers of Common Grounds. Elsevier, London, pp. 47-66.

Grice, H. Paul, 1989. Studies in the Way of Words. Harvard University Press, Cambridge, MA.

Haeseryn, Walter, Romijn, Kirsten, Geerts, Guido, de Rooij, Jaap, van den Toorn, Maarten Cornelis, 1997. Algemene Nederlandse Spraakkunst. Wolters Noordhoff, Groningen/Leuven.

Hall, Alison, 2004. The meaning of but: A procedural reanalysis. UCL Working Papers in Linguistics 16, 199-236.

Hall, Alison, 2007. Do discourse connectives encode concepts or procedures? Lingua 117 (1), 149-174.

Horn, Laurence R., 2004. Implicature. In: Laurence R. Horn & G. Ward (Eds.), The Handbook of Pragmatics. Blackwell, Oxford, pp. 3-28.

Huang, Yan, 2006. Pragmatics. Oxford University Press, New York.

Iten, Corinne, 2005. Linguistic Meaning, Truth Conditions and Relevance. The Case of Concessives. Palgrave Macmillan, New York.

Lepère, Jo , 2008. Het gebruik van 'maar' en 'toch' bij negenjarige kinderen. Een experimenteel onderzoek, Unpublished master's thesis. Catholic University of Leuven, Faculty of Arts, Department of Linguistics.

Noveck, Ira A., 2001. When children are more logical than adults: experimental investigations of scalar implicature. Cognition 78 (2), 165-188.

Schmerse, Daniel, Lieven, Elena, Tomasello, Michael, 2014. Discourse particles and belief reasoning: The case of German *doch*. Journal of Semantics 31, 115-133.

Sperber, Dan, Wilson, Deirdre, 1995. Relevance: Communication and Cognition. Blackwell, Oxford.

Van Belle, William, 2003. Zwijgen Is Niet Altijd Toestemmen. De Rol van Inferenties bij het Interpreteren en Argumenteren. Uitgeverij Acco, Leuven.

Van Belle, William, Devroy, Goedele, 1992. Tegenstellende en Toegevende Connectoren. Een Argumentatieve Beschrijving (Preprint 143). Catholic University of Leuven, Faculty of Arts, Department of Linguistics.

Wilson, Deirdre, Sperber, Dan, 1993. Linguistic form and relevance. Lingua 90, 1–25.

Wilson, Deirdre, Sperber, Dan, 2004. Relevance theory. In: Laurence R. Horn & G. Ward (Eds.), The Handbook of Pragmatics. Blackwell, Oxford, pp. 607-632.

APPENDIX A: MATERIALS EXPERIMENT 1 (translated from Dutch)

1. **Laurens has to find sponsors for an event he is organizing. He asks his neighbor, Thomas, if he wants to go along with him to search for sponsors in the neighborhood. Thomas is not sure whether or not to help Laurens.**

   Thomas thinks: "I live in a small house, but Laurens is a good friend."

   - "Nevertheless I will not join Laurens to find sponsors."
   - "Nevertheless I will join Laurens to find sponsors."

2. **Each year in the H.Hart Heverlee high school a teacher is awarded with the title of 'teacher of the year'. There is speculation among the students who will get the title this year. Bob, a boy from that school, wonders whether Mr. Van Damme will get the title.**

   Bob says: "Mr. Van Damme is a very good teacher, but he tends to favor certain students."

   - "Nevertheless I think he won't receive the title of 'teacher of the year'."
   - "Nevertheless I think he will receive the title of 'teacher of the year'."

3. **The junior soccer team of Haasrode plays a match against the junior soccer team of Bierbeek. Two spectators are discussing the potential chances of victory for the team from Haasrode. Mark doubts if the soccer team from Haasrode will win.**

   Mark says: "The soccer team of Haasrode is the best team in the ranking, but they have many injured players. "

   - "So they have little chance to win."
   - "So they have lots of chance to win."

4. **Santa Claus has brought many sweets this year. Evelien received lots of chocolate. She doubts whether she would eat chocolate or not.**

   Evelien thinks: "Chocolate is very tasty, but I have blond hair."

   - "So I won't eat chocolate."
   - "So I will eat chocolate."

5. **At school, the children learn about healthy food. Miss Els teaches about fruit and vegetables. She asks her class if bananas are healthy.**

   Gust answers: “There's a lot of sugar in bananas, but bananas contain lots of vitamins.”

   - “Nevertheless I think that bananas are not healthy.”
   - “Nevertheless I think that bananas are healthy.”

6. **During the holidays, Sarah is staying with her grandmother for a couple of days. Sarah is sitting outside in the garden when her grandmother brings her a glass of lemonade. Sarah is unsure whether or not to drink the lemonade.**

   Sarah says: “There are flowers in the garden, but I don’t like lemonade.”

   - “So I will not drink lemonade.”
   - “So I will drink lemonade.”

7. **Maggie, the cat of the Mertens family, gave birth to four kittens. The Mertens family wants to give the kittens to the Peeters family. The Peeters family is unsure if they will keep the kittens.**

   Mom says: “Dad is allergic to cats, but the kittens are black with white paws.”

   - “Nevertheless we will not keep the kittens.”
   - “Nevertheless we will keep the kittens.”

8. **Marie has bought a plant for Mom's birthday. She wants the plant to live as long as possible. Marie doubts whether this will succeed.**

   Marie thinks: “The plant gets no sunlight, but Mom has a big car.”

   - “So I am convinced that the plant won’t stay alive.”
   - “So I am convinced that the plant will stay alive.”

9. **Peter's friends decide to have a hamburger for lunch. Peter is unsure whether or not to join them.**

   Peter says: "I have curly hair, but I'm hungry."

   - "Nevertheless I will not join my friends for lunch."
   - "Nevertheless I will join my friends for lunch."

10. **Ellen goes to a concert tonight and asks Katrien if she would go with her. Katrien doubts whether she will go to the concert.**

    Katrien says to Ellen: "I have two fish, but the concert is expensive."

    - "Nevertheless I will not go to the concert."
    - "Nevertheless I will go to the concert."

11. **Mom and Jens are going to the dentist. Mom asks Robbe if he will come along. Robbe doubts whether he would go to the dentist or not.**

    Robbe says to Mom: "I'm afraid of the dentist, but I'm wearing a red sweater."

    - "So I will not go to the dentist."
    - "So I will go to the dentist."

12. **Miss Klara returns the exams that she corrected yesterday. Luke wonders if he will have a good grade.**

    Luke thinks: "I did not understand the subject matter, but I could solve all the questions."

    - "So I think I will not have a good grade."
    - "So I think I will have a good grade."

13. **The Verhulst family wants to spend an evening playing board games. Everyone participates. Only the oldest son, Alexander, is still unsure whether he will play or not.**

    Alexander thinks: "I like to take part in family activities, but I still have work to do."

    - "So I won't play board games with my family."
    - "So I will play board games with my family."

14. **It's Christmas. The De Corte family bought a Christmas tree and they want to decorate it. Mom is in doubt whether or not to let her youngest daughter, Sarah, help.**

    Mom explains: "Sarah is wearing pink pyjamas, but she likes decorating the tree."

    - "So Sarah cannot help."
    - "So Sarah can help."

15. **One day Sophie was at home playing with her ball. She knew she actually wasn't allowed to do that inside the house. When playing with the ball she broke a vase. When Mom comes home in the evening Sophie doubts if she will confess how the vase got broken.**

    Sophie says to her brother: "I want to be honest, but my favorite color is red."

    - "Nevertheless I will not confess how the vase got broken."
    - "Nevertheless I will confess how the vase got broken."

16. **Before Lennert and Vincent go to sleep Dad comes by to give them a goodnight kiss. The boys ask Dad if he wants to tell them a horror bedtime story. Dad doubts whether he would do so.**

    Dad says to Mom: "The boys have bunk beds, but they could get nightmares from a horror story."

    - "So I won't tell them a horror story."
    - "So I will tell them a horror story."

17. **Mom and Ella are shopping. Ella sees a lovely teddy bear lying on the shelves. She asks Mom if she can have the teddy bear. Mom is not sure.**

    Mom thinks: "Ella had a good report card, but it's Saturday."

    - "Nevertheless Ella cannot have the teddy bear."
    - "Nevertheless Ella can have the teddy bear."

18. Klaas is playing on the playground at school. He sees how Tim, his best friend, beats another child. Klaas doubts whether he should tell the teacher.

Klaas says: "Tim is my best friend, but I feel pity for the other child."

- "Nevertheless I will not tell the teacher what Tim did."
- "Nevertheless I will tell the teacher what Tim did."

APPENDIX B: MATERIALS EXPERIMENT 2 (translated from Dutch)

1. **Laurens has to find sponsors for an event he is organizing. He asks his neighbor, Thomas, if he wants to go along with him to search for sponsors in the neighborhood. Thomas is not sure whether or not to help Laurens.**

   Thomas thinks: "Laurens is a good friend, but I live in a small house."

   ➔ Will Thomas help Laurens to find sponsors?

   - o Yes
   - o No

   ➔ Why do you think that?

   ……

2. **Each year in the H.Hart Heverlee high school a teacher is awarded with the title of 'teacher of the year'. There is speculation among the students who will get the title this year. Bob, a boy from that school, wonders whether Mr. Van Damme will get the title.**

   Bob says: "Mr. Van Damme is married and has children, but he is a very good teacher."

   ➔ Will Mr. Van Damme get the title of 'teacher of the year'?

   - o Yes
   - o No

   ➔ Why do you think that?

   ……

3. **The junior soccer team of Haasrode plays a match against the junior soccer team of Bierbeek. Two spectators are discussing the potential chances of victory for the team from Haasrode. Mark doubts if the soccer team from Haasrode will win.**

   Mark says: "The soccer team of Haasrode is the best team in the ranking, but they have many injured players. "

   ➔ Will the soccer team from Haasrode win the match?

   - o Yes
   - o No

   ➔ Why do you think that?

   ……

4. **Santa Claus has brought many sweets this year. Evelien received lots of chocolate. She doubts whether she would eat chocolate or not.**

   Evelien thinks: "It's almost dinner time, but chocolate is very tasty."

   ➔ Will Evelien eat chocolate?

   - o Yes
   - o No

   ➔ Why do you think that?

   ……

5. **At school, the children learn about healthy food. Miss Els teaches about fruit and vegetables. She asks her class if bananas are healthy.**

   Gust answers: "Bananas contain lots of vitamins, but there's a lot of sugar in bananas."

   ➔ Does Gust think that bananas are healthy?

   - o Yes
   - o No

   ➔ Why do you think that?

   ……

6. **During the holidays, Sarah is staying with her grandmother for a couple of days. Sarah is sitting outside in the garden when her grandmother brings her a glass of lemonade. Sarah is unsure whether or not to drink the lemonade.**

   Sarah says: "I don't like lemonade, but there are flowers in the garden."

   ➔ Will Sarah drink the lemonade?

   - Yes
   - No

   ➔ Why do you think that?

   ……

7. **Maggie, the cat of the Mertens family, gave birth to four kittens. The Mertens family wants to give the kittens to the Peeters family. The Peeters family is unsure if they will keep the kittens.**

   Mom says: "The kittens are black with white paws, but Dad is allergic to cats."

   ➔ Will the Peeters family keep the kittens?

   - Yes
   - No

   ➔ Why do you think that?

   ……

8. **Marie has bought a plant for Mom's birthday. She wants the plant to live as long as possible. Marie doubts whether this will succeed.**

   Marie thinks: "The plant gets lots of water, but Mom has a big car."

   ➔ Will the plant live long?

   - Yes
   - No

   ➔ Why do you think that?

   ……

9. **Peter's friends decide to have a hamburger for lunch. Peter is unsure whether or not to join them.**

   Peter says: "I have curly hair, but I'm hungry."

   ➔ Will Peter join his friends for lunch?

   - Yes
   - No

   ➔ Why do you think that?

   ……

10. **Ellen goes to a concert tonight and asks Katrien if she would go with her. Katrien doubts whether she will go to the concert.**

    Katrien says to Ellen: "The concert is expensive, but I have two fish."

    ➔ Will Katrien go to the concert?

    - Yes
    - No

    ➔ Why do you think that?

    ……

11. **Mom and Jens are going to the dentist. Mom asks Robbe if he will come along. Robbe doubts whether he would go to the dentist or not.**

    Robbe says to Mom: "I'm afraid of the dentist, but I have a tooth ache."

    ➔ Will Robbe go to the dentist?

    - Yes
    - No

    ➔ Why do you think that?

    ……

12. Miss Klara returns the exams that she corrected yesterday. Luke wonders if he will have a good grade.

Luke thinks: "I came to school by bike, but I did not understand the subject matter."

➔ Does Luke think that he will have a good grade?

- Yes
- No

➔ Why do you think that?

……

13. The Verhulst family wants to spend an evening playing board games. Everyone joins in the game. Only the oldest son, Alexander, is still unsure whether he will play or not.

Alexander thinks: "I still have work to do, but I can sing beautifully."

➔ Will Alexander play along with the board games?

- Yes
- No

➔ Why do you think that?

……

14. It's Christmas. The De Corte family bought a Christmas tree and they want to decorate it. Mom is in doubt whether or not to let her youngest daughter, Sarah, help.

Mom explains: "Sarah likes decorating the tree, but she is very clumsy."

➔ Will Mom let Sarah help decorating the tree?

- Yes
- No

➔ Why do you think that?

……

15. **One day Sophie was at home playing with her ball. She knew she actually wasn't allowed to do that inside the house. When playing with the ball she broke a vase. When Mom comes home in the evening Sophie doubts if she will confess how the vase got broken.**

    Sophie says to her brother: "I want to be honest, but my favorite color is red."

    ➔ Will Sophie confess how the vase got broken?

    - Yes
    - No

    ➔ Why do you think that?

    ……

16. **Before Lennert and Vincent go to sleep Dad comes by to give them a goodnight kiss. The boys ask Dad if he wants to tell them a horror bedtime story. Dad doubts whether he would do so.**

    Dad says to Mom: "The boys have bunk beds, but they could get nightmares from a horror story."

    ➔ Will Dad tell his sons a horror story?

    - Yes
    - No

    ➔ Why do you think that?

    ……

17. **Mom and Ella are shopping. Ella sees a lovely teddy bear lying on the shelves. She asks Mom if she can have the teddy bear. Mom is not sure.**

    Mom thinks: "Ella has been bad, but it's Saturday."

    ➔ Can Ella have the teddy bear?

    - Yes
    - No

    ➔ Why do you think that?

    ……

18. Klaas is playing on the playground in school. He sees how Tim, his best friend, beats another child. Klaas doubts whether he should tell the teacher.

Klaas says: "I'm in love, but I feel pity for the other child."

➔ Will Klaas tell the teacher what Tim did?

- Yes
- No

➔ Why do you think that?

……

Chapter 4

But: do age and working memory influence conventional implicature processing?

Abstract

Conventional implicatures are omnipresent in daily life communication but experimental research on this topic is sparse, especially research with children. The aim of this study was to investigate if 8-to-12-year-old children spontaneously make the conventional implicature induced by *but, so* and *nevertheless* in 'p but q' sentences. Additionally, the study aimed to shed light on the cognitive effort required for these inferences by measuring working memory (WM) capacity. Our results show that children do make these inferences to a certain extent, but are sensitive to the content of the arguments. We found a significant effect of sentence type, but did not observe any developmental effect, nor any effect of WM: a higher age or WM capacity does not result in more pragmatic inferences.

Janssens, L., Drooghmans, S., & Schaeken, W. (in press). But: do age and working memory influence conventional implicature processing? *Journal of Child Language*.

Introduction

Within the field of pragmatics, a large body of experimental research has been devoted to implicatures. The majority of this research has focused on the category of conversational implicatures. A conversational implicature is "a component of speaker meaning that constitutes an aspect of what is meant in a speaker's utterance without being part of what is said" (Horn, 2004:3). A widely investigated subtype of conversational implicatures is the scalar implicature. The scalar implicature arising from a sentence like *some children are naughty* is that *some but not all children are naughty.* In contrast to this pragmatic interpretation, this utterance can also be interpreted logically as *some and perhaps all children are naughty*. However, the general assumption is that, whenever a weaker term (e.g., the quantifier *some*) is used, a stronger term from the same scale (e.g., *all*) does not hold, or that the speaker does not know whether it holds, because otherwise they would be underinformative. This interpretation can be applied to a range of different scales (e.g., <*always, often, sometimes*>, <*three, two, one*>, etc.).

In this paper, however, we will focus on the other category of implicatures, conventional implicatures, which have received much less attention in the literature. Unlike conversational implicatures, conventional implicatures are related to the conventional meaning of words. A typical example is that *but* conventionally implies a contrast. In an utterance such as *he is short, but he is healthy*, the use of *but* suggests that there is a contrast between 'short' and 'healthy', even though this contrast is not explicitly expressed.

Grice (1989) introduced the concept 'conventional implicature' to describe conventional aspects of meaning that are not truth-conditional. For example:

(1a) He is short but he is healthy.

(1b) He is short and he is healthy.

Truth-conditionally, (1a) and (1b) have the same meaning. However, *but* is not a truth-conditional connector because *but* provides an additional meaning to the utterance in (1a). It suggests that there is a contrast between the first and the second clause. This additional feature of *but* is what Grice describes as a conventional implicature. However, within the context of relevance theory, Blakemore (2002) proposed a procedural analysis of *but*, which can be summarized as follows:

> "To say that *but* means denial is to say that it encodes a constraint that triggers an inferential route involving *contradicting and eliminating* an assumption that is *manifest* in the *context*." (in Hall, 2004:220)

Thus, according to Blakemore, *but* encodes a specific procedure. This procedural analysis conflicts with Grice's conceptual analysis in which *but* encodes the concept 'contrast'.

The experiment discussed in this paper focuses on the conventional implicature induced by the conjunction *but*. This work builds on Janssens and Schaeken (2013), but focuses on children instead of adults. Janssens and Schaeken (2013) investigated 'p but q' utterances constructed as distancing contrastive connections. In a distancing contrast, *but* connects two parts of a complex speech act (Van Belle & Devroy, 1992) and the second part is disassociated from the first part, without denying what the latter expresses (Haeseryn, Romijn, Geerts, De Rooij, & Van den Toorn, 1997). In the 'p but q' construction, the speaker endorses that *p* is true (Van Belle, 2003). However, because *but* is used, the possible inference from *p* is cancelled in favor of the conclusion stemming from *q*. The *but* construction indicates that the *q*-argument should be considered more relevant than the inference made from *p*. For example:

(2) I really like chocolate, but it's almost dinner time.

In a context where someone is offered a piece of chocolate, the two arguments of (2) lead to opposite conclusions. The first part (*p*) of this sentence elicits the conclusion that the speaker will eat the piece of chocolate, whereas the *q*-argument yields the opposite conclusion, i.e. that the speaker will not eat the chocolate. In this construction, *but* not only indicates that there is a contrast between the two arguments, it also provides more weight to the *q*-argument. The conclusion from the *q*-argument is therefore the pragmatic conclusion that follows from *but*. In other words, by uttering (2) the pragmatic conclusion is that the speaker will not eat the chocolate. Note that when the arguments are reversed (i.e. *it's almost dinner time, but I really like chocolate*), the opposite conclusion will follow, i.e. that the speaker will eat the chocolate.

The conclusion that follows from the 'p but q' construction can be introduced by words like *so* or *nevertheless*. These two words elicit opposite conclusions. When the word *so* follows a 'p argumentative but q' utterance, it introduces the expected conclusion from *q* (e.g., *I really like chocolate, but it's almost dinner time. So I won't eat*

chocolate.). In contrast, Lepère (2008) argued that, according to Van Belle (2003), the word *nevertheless*[1], used as a conjunctive adverb, reverses the argumentative orientation of a 'p argumentative but q' sentence. It overrules the conclusion from *q* and redirects the reader towards the conclusion implied by *p* (e.g., *I really like chocolate, but it's almost dinner time. Nevertheless, I will eat chocolate.*). Note that the argumentative strength of *but*, *so* and *nevertheless* should be considered separate from the content of the arguments, which they are supposed to take precedence over (Van Belle, 2003).

Janssens and Schaeken (2013) presented adult participants with short stories, each ending with a 'p but q' sentence. The *p*- and *q*-arguments were either sensible or irrelevant and always led to opposite conclusions. For instance, 'I really like chocolate' and 'it's almost dinner time' are both sensible arguments (for eating and not eating chocolate, respectively). However, in this context – in which someone is offered a piece of chocolate – 'I have blonde hair' would be an example of an irrelevant argument. An irrelevant argument has no relation with the story and in no way indicates which conclusion should follow. Such irrelevant arguments were included in order to investigate whether people might be influenced by the content of the arguments rather than the structure of the sentence (i.e. the pragmatic meaning of *but*).

After participants read the 'p but q' sentence, they were asked to indicate the appropriate conclusion introduced by *so* ('*so* conclusion from *p*' or '*so* conclusion from *q*') for one half of the stories, and the one introduced by *nevertheless* ('*nevertheless* conclusion from *p*' or '*nevertheless* conclusion from *q*') for the other half. The 'appropriate' pragmatic conclusion introduced by *so* is the one stemming from the *q*-argument, while for *nevertheless* this is the one from the *p*-argument. The results showed that, in general, people do make the inference induced by *but*. This was clear from the items with two sensible arguments for which the *so*-conclusion was asked (82% pragmatic answers). In contrast, for the *nevertheless*-conclusions, the expected conclusions from *p* were given in only 48% of the cases. This could indicate that the meaning of *nevertheless* does not give as much precedence to the *p*-argument as was expected. On the other hand, this latter finding could also be explained by the fact that the inference stemming from *nevertheless* is opposite to the one from *but*. In a 'p but q'

[1] The experiment described in this paper was carried out in Dutch. Throughout this paper we use *nevertheless* to translate Dutch *toch*, even though these two adverbs do not have the exact same meaning.

construction, *but* leads the reader towards the conclusion from *q*, but *nevertheless* overrules this conclusion in favor of the conclusion from *p*. This understandably requires more effort than a *so*-conclusion, which explains why the number of appropriate answers was lower for *nevertheless*.

The content of the arguments was also found to play an important role. Whenever an irrelevant argument was combined with a sensible argument, most participants favored the conclusion from the sensible argument, regardless of the conventional meaning of *but*, *so* or *nevertheless*. The importance of the content was confirmed in a second experiment in which participants were asked to justify their answer. As expected, whenever their conclusion was not the pragmatic one, participants mostly explained this by referring to the content of the arguments.

This paper investigates whether the pragmatic meaning of *but*, combined with *so* and *nevertheless*, is also clear to children. In conversational implicature research, it has been shown that children are less able to make pragmatic inferences than adults. For example, Noveck (2001) found that 89% of the seven-to-eight-year-olds in his study agreed with statements such as *some giraffes have long necks*, compared to only 41% of the adults. Similarly, with respect to propositional connectives, Braine and Rumain (1981) presented evidence that deductively competent seven- and nine-year-old children favor a logical interpretation of *or* ('p or q and perhaps both') over an implicit one ('p or q but not both'). Adults given the same task were equivocal, although they tended to favor exclusive interpretations (Braine & Rumain, 1981). Such observations that children are less able to make pragmatic inferences than adults have led to further research into factors that enhance children's pragmatic competence. For example, in an experiment with five-year-old children, Papafragou and Musolino (2003) found that a training session prior to the presentation of the test sentences in order to enhance the children's awareness of pragmatic anomalies caused the number of pragmatic answers to increase.

For conventional implicatures, we also expect that children are less able to make pragmatic inferences than adults. This is because there is a certain similarity between conversational implicatures and the conventional implicatures investigated in our experiment. A specific feature of conventional implicatures – in contrast with conversational implicatures – is that they are not cancellable. However, *but* sentences seem to carry two implicatures. First, there is the implicature that *but* creates a contrast

between *p* and *q* (e.g., *he is short, but he is healthy*), which indeed seems to be a classic, non-cancellable conventional implicature. However, this is not the case for the second implicature, i.e. that *but* provides more weight to the *q*-argument (e.g., *I really like chocolate, but it's almost dinner time*). The use of *nevertheless*, for example, can lead to a cancellation of this implicature as it provides more weight to the *p*-argument in a 'p but q' utterance. Accordingly, the conventional implicatures that we investigate in this paper may not be purely conventional, but they share certain features with conversational implicatures. That is why we expect similar results for the conventional implicature task with *but*, i.e. that children are less able to make these pragmatic inferences.

The typical pragmatic development in children for conversational implicatures is often explained by the fact that drawing the implicature requires effort and children have less cognitive resources available than adults. Cognitive effort in adults has been tested by De Neys and Schaeken (2007), among others, who found that burdening working memory (WM) with a secondary task decreases pragmatic processing. Together with other observations (e.g., Bott & Noveck, 2004; Noveck & Posada, 2003), this was taken as evidence that scalar implicature processing is effortful. Since the conventional implicatures we investigate in this paper have certain properties in common with conversational implicatures (i.e. they are cancellable), we also examined whether WM plays a role in processing these implicatures. In addition to testing children's understanding of the pragmatic meaning of *but, so* and *nevertheless*, we also measured whether a higher WM span corresponds with a better understanding of the pragmatic meaning of *but, so* and *nevertheless.*

Adults had been found to clearly grasp the pragmatic meaning of *but* (Janssens & Schaeken, 2013). However, they also proved to be influenced by the content of the arguments, in spite of their pragmatic competence. For this reason, it is plausible to expect that children will definitely show great sensitivity to the content of the arguments in 'p but q' constructions. This expectation is based on the similarity between conversational implicatures and the conventional implicatures focused on in this paper, but this expectation can also be deduced from the findings of Schaeken, Sevenants and Madruga (2011), who observed a clear effect of content in children. For 9-to-13-year-old children, who were given a reasoning task with *unless*, abstract problems proved to be much more difficult to grasp than concrete problems. The concrete, meaningful problems were about daily life situations that could easily be imagined in a realistic

context (e.g., *you will have to go to bed early, unless you make no more mistakes*) whereas the abstract problems concerned cards with letters on one side and numbers on the other side (e.g., *there is a '2', unless there is not an 'A'*). This suggests that for children to be able to reason with *unless,* they need to be presented with a meaningful context that is imaginable in real life. Children seem to need meaning to reason. We therefore expect the children in our experiment to be greatly influenced by the type of sentence, and more specifically by the content of the arguments. We expect that when children see an argument that they judge as a very strong argument in its context, the content of the arguments will often prevail over the conventional meaning of *but*, *so* and *nevertheless*. Especially the combination of sensible arguments with irrelevant arguments is likely to make a difference. Furthermore, the interaction between the type of argument (sensible or irrelevant) and the type of conclusion (*so* or *nevertheless*) observed in adults can also be expected in children. We expect the irrelevant arguments to facilitate the *so*-conclusions in the 'irrelevant but sensible' contexts, and the *nevertheless*-conclusions in the 'sensible but irrelevant' contexts. Moreover, since adults were found to have great difficulty with *nevertheless*, we expect these conclusions to be especially difficult for children.

Method

Participants

A total of 86 children (39 boys and 47 girls) between 8 and 12 years old, with a mean age of 10;6, participated in this study. They were selected from two different schools in Belgium and were all native Dutch speakers.

Working Memory Task

In order to relate the results of the implicature task to the children's WM span, we measured WM by means of a Listening Span task (Daneman & Carpenter, 1980). In the Listening Span task, the children heard the experimenter read utterances aloud and were asked to write down whether these utterances were true or false. In addition, they had to remember the last word of every utterance and write these words down in the correct order at the end of each trial. They started with three trials with a listening span of one utterance. The span length was increased by one utterance whenever the children

wrote the words down in the correct order in at least two of the three trials. Every increase of the listening span was announced by the experimenter in order to reduce the effect of attentional factors. Whenever at least two out of three trials were written down incorrectly, the Listening Span task was terminated. A trial was scored as correct when all the words of this trial were written down in the correct order. The total score was the sum of all correct trials.

Implicature Task

The implicature task consisted of 18 context stories that were adopted from Janssens and Schaeken (2013). Each of the stories described a person in doubt about something. For example:

> *It's Christmas. The 'De Corte' family bought a Christmas tree and wants to decorate it. Mom is in doubt whether to let her youngest daughter Sarah help.*

After the short story, the person in doubt gives two contrastive arguments separated by *but*. For example:

> *Mom thinks: "Sarah likes decorating the tree, but she is very clumsy."*

The arguments were either sensible or irrelevant. In the example above, both arguments are sensible. These sensible arguments have an argumentative orientation determined by either a positive or negative value that we ascribe to its content, which is referred to as the 'axiological value' by Anscombre and Ducrot (1977). A positive argument (e.g., *Sarah likes decorating the tree*) is an argument of which the axiological value is oriented towards a positive conclusion (e.g., Sarah can help), while a negative argument (e.g., *she is very* clumsy) elicits a negative conclusion (e.g., Sarah cannot help).

The irrelevant arguments have no positive or negative axiological value, i.e. they are not oriented towards a positive or negative conclusion. An example of an irrelevant argument in this context is the following:

> *Mom thinks: "Sarah is very clumsy but she is wearing pink pyjamas."*

In total, there were six combinations of arguments: 'positive sensible but negative sensible', 'negative sensible but positive sensible', 'irrelevant but positive sensible', 'irrelevant but negative sensible', 'positive sensible but irrelevant' and 'negative sensible

but irrelevant'. There was no 'irrelevant but irrelevant' combination because if both arguments are unrelated to the context of the story, there is no way to know what the conclusion should be. See Table 1 for an example from the Christmas story of every sentence type.

Table 1 *Example of each sentence type*

Sentence type	**Example**
pos S but neg S	*Sarah likes decorating the tree, but she is very clumsy.*
neg S but pos S	*Sarah is very clumsy, but she likes decorating the tree.*
I but pos S	*Sarah is wearing pink pyjamas, but she likes decorating the tree.*
I but neg S	*Sarah is wearing pink pyjamas, but she is very clumsy.*
pos S but I	*Sarah likes decorating the tree, but she is wearing pink pyjamas.*
neg S but I	*Sarah is very clumsy, but she is wearing pink pyjamas.*

pos=positive, neg=negative
S=sensible, I=irrelevant

After participants read the 'p but q' construction, they were asked to choose the appropriate conclusion. For half the stories, they had to choose the appropriate conclusion introduced by *so*, and for the other half of the stories the appropriate conclusion introduced by *nevertheless*. For example:

1. *"So Sarah can help."*
2. *"So Sarah cannot help."*

or

1. *"Nevertheless Sarah can help."*
2. *"Nevertheless Sarah cannot help."*

The appropriate pragmatic conclusion with *so* is the conclusion that follows from the *q*-argument, whereas the expected appropriate conclusion with *nevertheless* is the conclusion resulting from the *p*-argument.

Procedure

First, the Listening Span task was conducted with groups of five or six children at the same time. This task was performed in their classroom at school. In order to familiarize the children with the task, they were given three training trials, followed by the real task. Then, all children completed the implicature task. This was an individual written task which all children performed simultaneously in their classroom at school. Each participant answered three items from each argument combination, with half the items consisting of *so*-questions and the other half of *nevertheless*-questions.

Results

In order to analyze whether children made the correct inferences, we made no distinction between positive and negative arguments. There is no reason to expect that one would be more difficult than the other, as was confirmed by the finding that separate analyses did not present any significant differences. Therefore, for our analyses, we collapsed the items to the combinations: 'sensible but sensible', 'irrelevant but sensible' and 'sensible but irrelevant'. Since children always had a one out of two chance of giving the expected answer, we also analyzed whether their performance differed significantly from chance level. The results, displayed in Table 2, show that performance on each sentence type differed significantly from chance level. Performance on the sentence types 'sensible but sensible, so', 'irrelevant but sensible, so' and 'sensible but irrelevant, nevertheless' was significantly above chance level and performance on all other sentence types was significantly below chance level ('sensible but irrelevant, so', 'sensible but sensible, nevertheless' and 'irrelevant but sensible, nevertheless').

Table 2 *Percentages of appropriate so- and nevertheless-conclusions for each sentence type*

Sentence type	**Percentage of appropriate answers (*N*=86)**
SS_so	64.54***
IS_so	80.04***
SI_so	31.10***
SS_nevertheless	40.69**
IS_nevertheless	29.84***
SI_nevertheless	75.88***

I=irrelevant, S=sensible
*p<.05, **p<.01, ***p<.001

Since participants were nested in different age groups and the dependent variable was binary, a generalized linear mixed model (also known as multilevel or hierarchical linear modeling) with a logit link function was used to analyze the data (see e.g., Baayen, Davidson, & Bates, 2008; Bates, Maechler, & Bolker, 2011; or Jaeger, 2008). The model fitting procedure was implemented in R using the lmer() function from the lme4 package. The model was made increasingly complex until model fit no longer increased, which was assessed using the Bayesian Information Criterion (BIC). The final model includes an effect of sentence type, but no main effects of age or WM capacity, nor interactions of these variables with sentence type.

In addition, we analyzed correlations between WM span and performance on the six sentence type categories. All correlations were non-significant, with no correlation above .096.

Table 3 displays an overview of the final model. The 'irrelevant but sensible, nevertheless' sentences are presented as the intercept with which all other sentence types are compared. We found that performance on these sentences did not differ significantly from 'sensible but irrelevant, so' nor from 'sensible but sensible, nevertheless'. The performance on all other sentence types did, however, differ significantly from 'irrelevant but sensible, nevertheless'.

Table 3 *Parameter estimates for the model with sentence type as a predictor*

Predictor	**Estimate**	**Standard Error**	***Z***	***p***
Intercept (IS_nevertheless)	-1.33	0.27	-5.01	<.001***
SI_nevertheless	2.88	0.39	7.42	<.001***
SS_nevertheless	0.38	0.36	1.06	0.29
SS_so	1.66	0.34	4.83	<.001***
SI_so	0.20	0.37	0.55	0.59
IS_so	2.52	0.37	6.86	<.001***

I=irrelevant, S=sensible
*p<.05, **p<.01, ***p<.001

Table 4 shows further pairwise comparisons for the different levels of sentence type. These were obtained by applying the multcomp package's glht() function on the final model. The reported p-values were adjusted for multiple comparisons using the single-step method. For the *so*-conclusions, the best results were obtained for the 'irrelevant but sensible' sentences. Performance on these sentences did not differ significantly from the 'sensible but sensible, so' sentences, but both sentence types did differ significantly from the 'sensible but irrelevant, so' sentences. For the *nevertheless*-conclusions, performance was very good on the 'sensible but irrelevant' sentences and the number of appropriate answers differed significantly from 'sensible but sensible, nevertheless' and 'irrelevant but sensible, nevertheless'. Finally, when we compare performance on *so*-conclusions with *nevertheless*-conclusions, the children assessed *so*-conclusions significantly more accurately than *nevertheless*-conclusions when presented with two sensible arguments.

Table 4 *Pairwise comparisons for the different sentence types*

Comparison	**Estimate**	**Standard Error**	***Z***	***p***
SS_so - IS_so	-0.87	0.34	-2.58	0.10
SI_so – IS_so	-2.32	0.36	-6.49	<.001***
SS_so – SI_so	1.46	0.33	4.38	<.001***
SS_so – SS_nevertheless	1.28	0.32	3.93	.001**
SS_nevertheless – SI_nevertheless	-2.50	0.37	-6.73	<.001***
SS_nevertheless – IS_nevertheless	0.38	0.36	1.06	0.90
IS_nevertheless – SI_nevertheless	2.88	0.39	7.42	<.001***

I=irrelevant, S=sensible
*p<.05, **p<.01, ***p<.001

General discussion

The aim of this paper was to determine whether children make the conventional implicature induced by *but* in 'p but q' constructions combined with *so* and *nevertheless*. While this had already been investigated in adults, no data were available on children's abilities to make these inferences. A group of 8-to-12-year-old children took part in our experiment. Furthermore, we wanted to address the question whether conventional implicature production is affected by WM span. Given their similarities with conversational implicatures, we examined whether processing these implicatures is effortful. To that end, we measured WM span by means of a Listening Span task.

The data were analyzed using a generalized linear mixed model, which showed that there was a significant main effect of sentence type, but no significant main effect of age or WM[2], nor any interaction between the two. When analyzing how the children performed compared to chance (see Table 2), we see that all results differed significantly from chance level. The children performed significantly below chance level on most of the sentences with *nevertheless.* Only the 'sensible but irrelevant, nevertheless' sentences were answered fairly accurately. This is because the appropriate answer for these items corresponds with the sensible argument, while the

[2] As could be expected, age and WM correlated significantly. According to Gathercole (1999) WM still improves through development and only reaches asymptotic levels at age 12.

irrelevant argument can be ignored. On the 'sensible but irrelevant, so' sentences, children also scored significantly lower than chance level. In this case following the sensible argument (and ignoring the irrelevant one) does not yield the appropriate answer. Given that all results differed significantly from chance level, we can conclude that children purposefully chose the inappropriate answer because they believed it to be the appropriate one, not because they were guessing. The results above chance level performance suggest that the meaning of *but* was fairly clear to the children, but they had difficulty grasping the meaning of *nevertheless*. This explains why they mostly chose the conclusion from *q*, to which they were directed by *but*, and not the appropriate conclusion from *p*. Moreover, this suggests that the content of the arguments is a very important factor, explaining the poor performance on 'sensible but irrelevant, so' and 'irrelevant but sensible, nevertheless'.

In general, our results showed that children are able to make the inference induced by *but* to a certain extent, but the percentages are rather low. Compared with the results obtained by Janssens and Schaeken (2013), the percentages pragmatic answers in children, given the exact same task, were clearly lower than in adults. This suggests that children's pragmatic understanding of *but* is not yet fully developed. Since even adults were found to be influenced by the content of the arguments, these results also indicate that this has even more of an impact on children. By comparing the performance on the different sentence types, this was made very clear. In the 'p but q' constructions that contained an irrelevant argument, children mainly based their answers on the sensible argument, irrespective of the conventional meaning of *but*, *so* or *nevertheless*. However, not only the irrelevant situations provided evidence that children are sensitive to content. Whenever two sensible arguments were presented and the *so*-conclusion was asked, children only provided 65% pragmatic answers (compared to 82% for the adults). This seems to indicate that their answer is often based on the argument they themselves deem most plausible and not necessarily on the *q*-argument.

Another parallel with the results of the adults is that the percentage of expected *nevertheless*-conclusions from the *p*-argument (following two sensible arguments) is very low: only 41%. As was argued in Janssens and Schaeken (2013), this could indicate that the assumed meaning of *nevertheless*, as reversing the expected conclusion from *but*, might not be correct. However, these results might also be explained by the additional effort required to make the correct interpretation. Since *nevertheless* requires

the reader to first make the inference from *but*, and then overrule this inference to draw the conclusion from the *p*-argument, it seems likely that this whole process might be more effortful than simply drawing the *so*-conclusion.

In order to determine whether processing the conventional implicature with *but* is effortful, we related the results of the implicature task to a WM test. Surprisingly, we found no effect of WM span, nor an effect of age. We expected these factors to play a significant role based on the similarity of this specific conventional implicature with conversational implicatures. It is important to mention that the absence of a WM effect cannot be attributed to the WM task. The WM scores ranged between 4 and 15 with a standard deviation of 2.52. This means that there was enough variability to identify a possible effect. The lack of an effect of WM (or age) may suggest that, in contrast to conversational implicatures, processing this implicature happens automatically and requires no WM. However, before drawing such a strong conclusion, we have to keep in mind that the effect of WM for conversational implicatures, although significant, is only small (e.g., see De Neys & Schaeken, 2007). WM involvement must be small in order to ensure smooth communication. Furthermore, when Dieussaert, Verkerk, Gillard and Schaeken (2011) found an effect of WM, this was only observed in participants with a lower WM capacity: burdening WM while performing a conversational implicature task had no effect on participants with a high WM capacity. Finally, while a significant effect of WM was observed in adults, Janssens and Schaeken (2012) found no significant WM effect on children's understanding of conversational implicatures.

Taking these considerations together, we must be cautious in making strong claims about the role of WM in conventional implicature processing. Moreover, the results obtained in this study reveal that children's understanding of the pragmatic meaning of *but*, *so* and *nevertheless* is not optimal yet and still strongly determined by the content of the arguments. This could indicate that the investigated age group may have been too young (i.e. insufficiently capable of this task) to reveal significant differences based on age or WM. Therefore, it might be better to focus on older age groups in future experiments, or make a direct comparison between children and adults. Furthermore, a different approach might be considered to investigate the role of WM. As in De Neys and Schaeken (2007), a double task design could be used, in which WM is burdened with a secondary task while performing the implicature task. A decrease of

pragmatic answers under WM load would indicate that conventional implicature processing does require WM involvement.

In sum, this study's main conclusion is that children's pragmatic understanding of *but* is not yet fully developed. Although they do seem to grasp its meaning to a certain extent, the content of the arguments has a very strong influence on their answers. Whenever one of the arguments is clearly more plausible than the other, the meaning of *but* seems to be of no importance.

References

Anscombre, J. & Ducrot, O. (1977). Deux 'mais' en français? *Lingua* **43**, 23–40.

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**, 390-412.

Bates, D., Maechler, M. & Bolker, B., (2011). lme4: Linear mixed-effects models using S4 classes, R package version 0.999375-42 http://www.R-project.org/

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers.* Cambridge: Cambridge University Press.

Bott, L. & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* **51**, 437-457.

Braine, M. & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology* **31**, 46-70.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* **19**, 450-466.

De Neys, W. & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology* **54**, 128-133.

Dieussaert, K., Verkerk, S., Gillard, E. & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology* **64**, 2352-2367.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences* **3**, 410-419.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J. & van den Toorn, M. C. (1997). *Algemene Nederlandse spraakkunst*. Groningen/Leuven: Wolters Noordhoff.

Hall, A. (2004). The meaning of but: A procedural reanalysis. *UCL Working Papers in Linguistics* **16**, 199-236.

Horn, L. R. (2004). Implicature. In L. R. Horn and G. Ward (Eds.), *The handbook of pragmatics,* 3-28. Oxford: Blackwell.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language* **59**, 434-446.

Janssens, L. & Schaeken, W. (2012). The role of task characteristics in children's scalar implicature production. *Proceedings of the Annual Conference of the Cognitive Science Society*. Cogsci. Sapporo, Japan, 509-514.

Janssens, L. & Schaeken, W. (2013). 'But' how do we reason with it: An experimental investigation of the implicature stemming from 'but'. *Journal of Pragmatics* **57,** 194-209.

Lepère, J. (2008). *Het gebruik van 'maar' en 'toch' bij negenjarige kinderen. Een experimenteel onderzoek.* Unpublished master's thesis. Catholic University of Leuven, Faculty of Arts, Department of Linguistics.

Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* **78**, 165-188.

Noveck, I. A. & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language* **85**, 203-210.

Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition* **86**, 253-282.

Schaeken, W., Sevenants, A. & Madruga, J. (2011). The Dramatic Effect of Content on Children's Unless Reasoning: Pragmatic Modulation or Reconstruction? *Proceedings of the Annual Conference of the Cognitive Science Society.* Cogsci. Boston, US, 1170-1175.

Van Belle, W. (2003). *Zwijgen is niet altijd toestemmen. De rol van inferenties bij het interpreteren en argumenteren.* Leuven: Uitgeverij Acco.

Van Belle, W. & Devroy, G. (1992). *Tegenstellende en toegevende connectoren. Een argumentatieve beschrijving* (Preprint 143, voorlopige publicatie). Catholic University of Leuven, Faculty of Arts, Department of Linguistics.

Chapter 5

‘But’ how do children judge it on a scale?

Abstract

This study examined children’s comprehension of the conventional implicature induced by *but*, combined with *so* and *nevertheless*, in ‘p but q’ sentences constructed as distancing contrastive connections. Based on the Pragmatic Tolerance Hypothesis of Katsos and Bishop (2011), a three-point scale was used as response format. Using a scale instead of a binary judgment task can reveal more insight in which factors are considered most important when processing ‘p but q’ sentences. The results indicated that the content of the *p*- and *q*-arguments plays a very important role when children process ‘p but q’ sentences. However, their use of the three-point scale also indicated that they are sensitive to the pragmatic meaning of *but*, *so* and *nevertheless*. These results must be interpreted cautiously since the children seemed to use the middle value on the scale around 30% of the time in each sentence category, which was not in line with our predictions. This might indicate that children experience a general incomprehension with this type of sentences and answer with the middle value on the scale because they simply don’t know the answer.

***Keywords**:* conventional implicature; *but*; scale; content

Janssens, L., Van den Broeck, A., & Schaeken, W. (in press). ‘But’ how do children judge it on a scale? *Proceedings of the Annual Conference of the Cognitive Science Society*. Cogsci. Quebec City, Canada, 23-26 July 2014.

Introduction

Over the past few decades, considerable experimental research has been devoted to scalar implicatures. Grice introduced the term *implicature* in the 1967 William James lectures to offer an explanation for how it is possible that an utterance can mean more than what is literally said. Scalar implicatures are a subcategory of conversational implicatures and are based on a scale of informativity. For example, on the scale <*all, most, many, some*> the use of the more informative *all* logically entails that *some* is also true. However, in an utterance such as *'Some Belgians like to drink beer'*, the pragmatic meaning of *some* causes the hearer to interpret this utterance as *'Some but not all Belgians like to drink beer'* even though the logical meaning of *some* is *'some and perhaps all'.* According to Grice (1989), people follow a set of maxims in communication in order to understand each other correctly. That's why the consensus applies that whenever a speaker uses a weak term such as *some*, a stronger term such as *all* does not hold. The speaker would not have been optimally informative if a stronger term applied.

Developmental conversational implicature research has shown that children are less pragmatic than adults. For example, Noveck (2001) found that 89% of the seven-to-eight-year-olds in his study agreed with statements such as *'Some giraffes have long necks'*, compared to only 41% of the adults. Similarly, with respect to propositional connectives, Braine and Rumain (1981) presented evidence showing that deductively competent seven- and nine-year-old children favor a logical interpretation of *or* (*'p or q and perhaps both'*) over an implicit one (*'p or q but not both'*). Adults on the same task were equivocal, though they tended to favor exclusive interpretations (Braine & Rumain, 1981). However, these (and other) studies claiming that children lack pragmatic competence have been criticized by Katsos and Bishop (2011). In their implicature studies, Katsos and Bishop (2011) argued that earlier studies mostly employed tasks that cannot differentiate between actual implicature derivation and mere sensitivity to violations of informativeness. The majority of studies concluding that children are more logical than adults used binary judgment tasks in which participants were instructed to judge an utterance as 'true' or 'false'. Katsos and Bishop (2011) argued that children might not reject underinformative sentences because they are tolerant to violations of informativeness. However, this doesn't mean that they are not sensitive to these violations. In order to test this Pragmatic Tolerance Hypothesis, Katsos and Bishop (2011, Experiment 2) instructed their participants to judge on a ternary scale how well a

fictional character described certain situations. They found that children's performance did not differ from adults'. Underinformative utterances were judged by both groups with the middle value on the scale. This shows that children understand that using for example *some*, when *all* would have been a more informative description, is not optimal. However, in a binary judgment task they would not penalize such a description as false whereas adults would. In previous research (e.g., Noveck, 2001) this falsely led to the conclusion that children lack pragmatic competence.

Besides conversational implicatures, Grice (1989) also distinguished the category of conventional implicatures. This paper will deal with this far less investigated category of conventional implicatures. Unlike conversational implicatures conventional implicatures (a) are related to the conventional meaning of words, (b) are immediate conclusions from utterances, (c) cannot be cancelled, and (d) are related to the form of an utterance, not the content. In an utterance such as *'He's old but he's smart'*, *but* conventionally implies a contrast. The use of *but* elicits the inference that 'old' and 'smart' contrast each other even though this is not explicitly expressed.

The experiments in this paper focus on the conventional implicature induced by the conjunction *but*. Our experiments build further on Janssens and Schaeken (2013) and Janssens, Drooghmans and Schaeken (in press). However, because of the important findings of Katsos and Bishop (2011) concerning conversational implicatures, we will apply a ternary scale instead of a binary judgment task. This allows us to test certain predictions about children's understanding of this conventional implicature that cannot be discovered by using a binary judgment task.

In Janssens and Schaeken (2013), 'p but q' utterances, constructed as indirect distancing contrastive connections, were examined. In a distancing contrast, *but* connects two parts of a complex speech act (Van Belle & Devroy, 1992) and the second part is disassociated from the first part, without denying what is being expressed in the first part (Haeseryn et al., 1997). For example:

(1) Hannah: "I really like these beautiful earrings, but they are very expensive."

In a 'p but q' construction, the speaker endorses that *p* is true (Van Belle, 2003). However, because *but* is used, the inference from the *p*-argument is cancelled in favor of the inference from the *q*-argument. In (1), the *p*-argument elicits the conclusion that Hannah will buy the earrings whereas the *q*-argument elicits the opposite conclusion

that she will not buy the earrings. The conventional meaning of *but* causes the *q*-argument to outweigh the *p*-argument so the appropriate conclusion from a 'p but q' sentence is inferred from the *q*-argument. Consequently, from (1), the conclusion follows that Hannah will not buy the earrings. If the two arguments trade places ('*they are very expensive, but I really like these beautiful earrings'*) the opposite conclusion will be drawn that Hannah will buy the earrings. This shows that the conventional meaning of *but* provides more weight to the *q*-argument irrespective of the content of the arguments. According to Anscombre and Ducrot (1977), every argument is determined by a certain positive or negative value ascribed to its content, which they labelled the 'axiological value'. The axiological value we ascribe to the arguments of an utterance is dependent on cultural specific common sense views (Van Belle & Devroy, 1992). In (1), the *p*-argument is oriented towards a positive conclusion (Hannah will buy the earrings) and the *q*-argument is oriented towards a negative conclusion (Hannah will not buy the earrings). That's why we label the *p*-argument in (1) as the positive argument and the *q*-argument as the negative argument.

The conclusion from a 'p but q' construction can be introduced by words such as *so* or *nevertheless*. The pragmatic meaning of these two words leads to opposite conclusions. *So* elicits the conclusion from *q* and therefore confirms the expected conclusion inferred from the pragmatic meaning of *but* (*I really like these beautiful earrings, but they are very expensive. So I will not buy them.*). In contrast, according to Van Belle (2003), whenever *nevertheless* -used as a conjunctive adverb- follows a 'p but q' sentence, it reverses the argumentative orientation again. The expected conclusion from *q* is overruled and the reader is redirected towards the conclusion inferred from *p* (*I really like these beautiful earrings, but they are very expensive. Nevertheless I will buy them)*. Note that *nevertheless* is used here as a translation of Dutch *toch*.

The adult participants in Janssens and Schaeken (2013) were presented with short stories that ended with a 'p but q' sentence. Both sensible (Se) and irrelevant (Ir) arguments were administered. In (1), both arguments are sensible in a context in which a woman is standing in a jewelry store. In this same context, uttering *'I really like these beautiful earrings, but I like spaghetti'* clearly contains an irrelevant *q*-argument. The irrelevant arguments were unrelated to the context of the stories and their purpose was to examine whether the pragmatic meaning of *but* is understood irrespective of the content of the arguments.

Each 'p but q' sentence was followed by two possible *so*-conclusions ('*so* conclusion from *p*' and '*so* conclusion from *q*') or by two *nevertheless*-conclusions ('*nevertheless* conclusion from *p*' and '*nevertheless* conclusion from *q*'). The participants were instructed to indicate the appropriate conclusion. Janssens and Schaeken (2013) expected the appropriate pragmatic conclusion following *so* to be the conclusion inferred from *q* and the appropriate conclusion following *nevertheless* to be the inferred conclusion from *p*. The general outline of the results showed that adults understand the pragmatic meaning of *but*. However, the content of the arguments plays a non-negligible role. Whenever an irrelevant argument was combined with a sensible argument, the participants practically always inferred the conclusion from the sensible argument, irrespective of the pragmatic inference from *but*, *so* and *nevertheless*. The importance of the content was confirmed in a second experiment in which participants were asked to justify their answer. As expected, participants mostly referred to the content of the arguments whenever they did not provide the appropriate conclusion. More evidence showing the importance of the content was found in the fact that 82% appropriate *so*-conclusions were given when two sensible arguments were presented. This means that 18% of the answers was based on the inappropriate *p*-argument which the participants probably judged as a better argument than *q*. Another finding was that inferring the appropriate *nevertheless*-conclusion is a lot more difficult than inferring the appropriate *so*-conclusion. Only 48% appropriate *nevertheless*-conclusions were given when two sensible arguments were presented. This could indicate that the expected pragmatic meaning of *nevertheless* might not be the right one. On the other hand, this finding can be explained by the fact that the inference stemming from *nevertheless* is opposite to the inference stemming from *but*. In order to make the appropriate inference from *nevertheless*, the inferred conclusion from *but* has to be cancelled. It seems plausible that this would require effort and therefore leads to a higher percentage of inappropriate answers.

Janssens et al. (in press) performed the same experiment as Janssens and Schaeken (2013) but with children aged 8 to 12. Additionally, they measured working memory (WM) in order to see whether WM is involved in processing the conventional implicature stemming from *but*. The children's results showed the same pattern as the adult data but the percentages of appropriate answers were lower. Moreover, no significant effect of WM was found.

In this paper we apply the methodology of Katsos and Bishop (2011) on children between the ages of 8 and 12. If children understand the pragmatic meaning of *but*, *so* and *nevertheless* but are also sensitive to the content of the arguments, we expect them to choose the middle value on the scale when they have to judge the appropriateness of a conclusion from a 'p but q' construction in which there is a conflict between the pragmatic answer and the answer based on the content. Since both the content and the conventional meaning of *but* can play a role in judging conclusions from 'p but q' sentences, different predictions can be made for each of the categories. A schematic view is presented in Table 1. This table depicts which of the two arguments (*p* or *q*) gets most weight based on (1) the content, (2) *but* and (3) the conclusion word (*so* or *nevertheless*). We can see in Table 1 that the content as well as *but* and the conclusion word guide the reader towards the conclusion from *q* in the 'IrSe *so*' sentences. That's why we predict a lot of optimal answers on the scale and no neutral (middle) answers. If the content is very important for children and they are rather tolerant with respect to *but*, then we also expect very few neutral answers for the 'SeIr *nevertheless*' sentences. When both arguments are sensible, the content should not play a role. When these sentences are combined with *nevertheless*, then *but* and *nevertheless* lead to opposite conclusions. This might lead to doubt, but also to inappropriate answers, depending on which of the two factors is more important. If children are not at all sensitive to the pragmatic meaning of *but*, *so* and *nevertheless*, we would expect many neutral answers for both the *so*- and the *nevertheless*- conclusions.

Table 1 *Indication of which argument has more weight for every sentence category*

Sentence category	Content	But	Conclusion word
SeSe_So	=	q	q
SeSe_Nevertheless	=	q	p
IrSe_So	q	q	q
IrSe_Nevertheless	q	q	p
SeIr_So	p	q	q
SeIr_Nevertheless	p	q	p

Se=sensible; Ir=irrelevant

Also, if children truly lack this sensitivity, we would expect no neutral answers for the 'IrSe *nevertheless*'- and the 'SeIr *so*' sentences. In both cases, the content guides them towards the inappropriate conclusion and this would not be corrected by *but*, *so* or *nevertheless*.

Experiment 1

Method

Participants

Sixty-six Dutch speaking children (31 boys and 35 girls) between the ages of 8 and 10 years with a mean age of 9.1 participated in this study. They were recruited from five classes of two different schools.

Implicature Task

The implicature task was based on Janssens and Schaeken (2013) but the design was adapted. The children were presented with 24 context stories. Each of the stories described a person in doubt about something. For example:

> *Peter's best friend is flying to Egypt to go on a diving holiday. He asks if Peter wants to come along. Peter is in doubt whether he will join his best friend or not.*

Each short story was followed by a 'p but q' sentence with two contrastive arguments expressing doubt. For example:

> *Peter thinks: "I'm afraid of flying, but I would like to learn how to dive."*

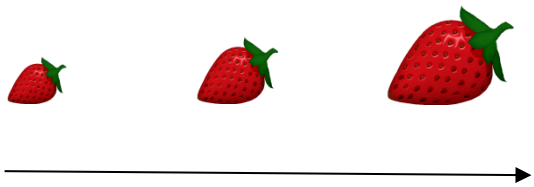
In the example above, both the *p*– and the *q*-argument are sensible arguments. The *p*-argument in this example is the negative argument (leading towards the negative conclusion '*I will not join my best friend on his trip')* and the *q*-argument is the positive argument ('*I will join my best friend on his trip'*). However, as in Janssens and Schaeken (2013), we also included irrelevant arguments in this experiment. The irrelevant arguments are not oriented towards a certain positive or negative conclusion. If the children understand the pragmatic meaning of *but* then these irrelevant arguments acquire a certain axiological value simply because they are contrasted with another

(positive or negative) argument. An example of a combination of an irrelevant and a positive argument is:

Peter thinks: "I like to eat chicken, but I would like to learn how to dive."

After each 'p but q' sentence, a certain conclusion was expressed by a fictional character, 'Mr. Coleman'. This conclusion was introduced by either *so* or *nevertheless*. Whenever *so* follows a 'p but q' sentence we expect the conclusion from the *q*-argument but when *nevertheless* follows, we expect the conclusion from *p*. The conclusions that were presented could be of four different kinds: '*so* conclusion from *p*', '*so* conclusion from *q*', '*nevertheless* conclusion from *p*' or '*nevertheless* conclusion from *q*'. After a certain conclusion was expressed, the children had to indicate on a scale how appropriate they judged the conclusion. Based on Katsos and Bishop (2011), we used a three-point scale with different sized strawberries. The children were instructed to reward a good conclusion with the biggest strawberry, a bad conclusion with the smallest strawberry and a conclusion that was neither completely bad nor good with the medium sized strawberry. E.g.:

Mr. Coleman says: "So Peter will join his best friend on his trip to Egypt."



The 24 stories represented an item from every combination of our 3x2x4 design. There were three possible argument combinations (SeSe, SeIr, IrSe), two different axiological value combinations (negative-positive, positive-negative) and four conclusion types ('*so q*', '*so p*', '*nevertheless q*', '*nevertheless p*').

Procedure

The task was administered to the children as a pen-and-paper task which they performed individually in their classroom at school. The task was introduced by a preliminary story about Mister Coleman who comes from America and wants to learn Dutch. The children were told that Mr. Coleman would utter several conclusions based

on each story and that they had to reward Mr. Coleman with different sized strawberries, depending on how appropriate the uttered conclusion was.

Results

In the analyses, we did not make a distinction between positive and negative arguments. When analyzing them separately, we did not find significant differences. That's why, in our analyses, we collapsed them. We recoded the children's answers into appropriate (three points), neutral (two points) and inappropriate (one point) answers. First we looked at the percentages of neutral answers for each of the argument conclusion combinations. These percentages are displayed in Table 2, together with the percentages of appropriate answers. There were no significant differences in the number of neutral answers between the different categories (X^2=1.21; *df*=2, *p*=.55). This is not in line with our expectations. We expected almost no neutral answers for the 'IrSe *so*' sentences and the 'SeIr *nevertheless*' sentences. However, since the number of neutral answers was evenly distributed over the different categories, this allowed us to sum up the scores in every category.

Table 2 *Percentages of neutral and appropriate (between brackets) answers for each argument conclusion combination (Experiment 1)*

Sentence	So	Nevertheless
Sensible-Sensible	31(41)	25(41)
Sensible-Irrelevant	29(25)	30(47)
Irrelevant-Sensible	28(55)	25(18)

When we look at the results of the *so*-conclusions, the children scored highest on the appropriateness scale for the 'IrSe' sentences (79%). This differed significantly from 'SeSe' (71%; Wilcoxon Signed Ranks test, *n*=55; *Z*=-3.66; *p*<.001) and from 'SeIr' (59%; Wilcoxon Signed Ranks test, *n*=66; *Z*=-5.28; *p*<.001). These last two categories also differed significantly from each other (Wilcoxon Signed Ranks test, *n*=56; *Z*=-4.18; *p*<.001). When we look at the results of the *nevertheless*-conclusions, the children scored highest on the appropriateness scale for the 'SeIr' sentences (75%). This differed significantly from the 'IrSe' sentences (54%; Wilcoxon Signed Ranks test, *n*=58; *Z*=-5.32;

p<.001) and marginally significantly from the 'SeSe' sentences (69%; Wilcoxon Signed Ranks test, *n*=54; *Z*=-1.89; *p*=.059). These last two categories also differed significantly from each other (Wilcoxon Signed Ranks test, *n*=53; *Z*=-5.18; *p*<.001).

In order to compare *so*-conclusions with *nevertheless*-conclusions, we have to look at the 'SeSe' sentences. We found no significant difference in performance between these two categories (Wilcoxon Signed Ranks test, *n*=49; *Z*=-.93; *p*=.35).

Discussion

The general outline of the results of the 8-to-10-year-olds seems to be in line with previous findings in Janssens et al. (in press). However, the introduction of a three-point scale enabled us to examine children's sensitivity to *but* in another way. The fact that children provide an inappropriate answer about half the time for the 'SeIr *so*'- and the 'IrSe *nevertheless*' sentences means that they provide an appropriate or neutral answer the other half of the time. As a consequence, this implies that, despite the importance of the content of the arguments, children are clearly sensitive to the pragmatic meaning of *but* and the conclusion words. However, Table 2 shows that the percentage of neutral answers is around 30% in each category. This is contrary to our expectations since we expected practically no neutral answers for the 'IrSe *so*' sentences. Because in these sentences not only the content, but also *but* and *so* guide the reader towards the conclusion from *q*, it is surprising that so many neutral answers were provided. This might point out that children could experience a general feeling of incomprehension and therefore prefer the middle value on the scale. Therefore we investigated slightly older children, aged 10 to 12, in Experiment 2. Perhaps a more clear answer pattern might emerge in older children. After all, childhood can be seen as a time where major changes are present in the development of different areas such as language, pragmatic- and logical understanding (Berk, 2010). We wondered whether there would be an age effect: will the older children in this experiment be more pragmatic than the younger children in Experiment 1 and will their use of the scale provide a clearer image of their understanding of *but*, *so* and *nevertheless*?

Experiment 2

Method

Participants, Materials and Procedure

The 61 Dutch speaking children who participated in this experiment were aged 10 to 12 with a mean age of 11.3. Two participants were excluded from the analyses due to missing data. The remaining children were 36 boys and 23 girls. They were students from the same schools as the children in Experiment 1 and were recruited from four different classes.

All materials and the procedure were exactly the same as in Experiment 1.

Results

The results of the older children are similar to those of the younger children in Experiment 1. We inspected the distribution of the neutral answers to see if it was permitted to sum up the scores. The percentages of the number of neutral answers are displayed in Table 3, together with the percentages appropriate answers. As in Experiment 1 there was no significant difference in the number of neutral answers between the different categories (X^2=.66; *df*=2, *p*=.72). This allowed us to sum up the scores in Experiment 2 as well and perform the same analyses as in Experiment 1.

When we look at the results of the *so*-conclusions, the 10-to-12-year olds scored highest on the appropriateness scale of the 'IrSe' sentences (87%). This differed significantly from the 'SeSe' sentences (79%; Wilcoxon Signed Ranks test, *n*=45; *Z*=-3.83; *p*<.001) and from the 'SeIr' sentences (57%; Wilcoxon Signed Ranks test, *n*=56; *Z*=-6.13; *p*<.001). These last two categories also differed significantly from each other (Wilcoxon Signed Ranks test, *n*=52; *Z*=-5.65; *p*<.001).

When we look at the results of the *nevertheless*-conclusions, the same pattern emerges as in Experiment 1. The children scored highest on the appropriateness scale of the 'SeIr' sentences (78%). This differed significantly from the 'SeSe' sentences (67%; Wilcoxon Signed Ranks test, *n*=59; *Z*=-3.51; *p*<.001) and from the 'IrSe' sentences (49%; Wilcoxon Signed Ranks test, *n*=54; *Z*=-5.99; *p*<.001). These last two categories also differed significantly from each other (Wilcoxon Signed Ranks test, *n*=51; *Z*=-5.06; *p*<.001).

Table 3 *Percentages of neutral and appropriate (between brackets) answers for each argument conclusion combination (Experiment 2)*

Sentence	So	Nevertheless
Sensible-Sensible	31(53)	32(35)
Sensible-Irrelevant	23(24)	21(57)
Irrelevant-Sensible	22(70)	19(14)

Table 4 *Results of Mann-Whitney U tests and percentages for the comparison between the two age groups*

Sentence	*U*	*p*	8-10 years	10-12 years
SeSe_So	1365.5	.003	71	79
IrSe_So	1329.5	.002	79	87
SeIr_So	1778.5	.400	59	57
SeSe_Nevertheless	1783.5	.410	69	67
IrSe_Nevertheless	1616	.096	54	49
SeIr_Nevertheless	1676.5	.180	75	78

Se=sensible; Ir=irrelevant

In contrast to Experiment 1, we did find a significant difference when we compared *so* with *nevertheless* for the 'SeSe' sentences (79% vs. 67% respectively; Wilcoxon Signed Ranks test, n =55; Z =-3.78; p <.001).

We performed Mann-Whitney U tests in order to explore the difference in performance between the two age groups. Generally, the older children provided more appropriate answers than the younger children but this difference was only significant for the 'SeSe *so*'– and the 'IrSe *so*' sentences. The results of these Mann-Whitney U tests are displayed in Table 4.

General Discussion

This paper aimed to examine children's understanding of the conventional implicature stemming from *but*, *so* and *nevertheless* in 'p but q' sentences constructed as distancing contrasts. Instead of using a binary judgment task as in Janssens et al. (in press), a three-point scale was used. The use of a ternary response format was inspired by Katsos and

Bishop (2011) who provided evidence that binary judgment tasks can conceal children's pragmatic competence. In line with their Pragmatic Tolerance Hypothesis they showed that children are equally aware of pragmatic violations as adults but are more tolerant for these violations. We expected a three-point scale to shed light on children's understanding of conventional implicatures as well. The results of Janssens et al. (in press) had shown that children seem to have a general understanding of the pragmatic meaning of *but*, *so* and *nevertheless* but are very sensitive to the content of the arguments. The use of a three-point scale enables the children to answer with the middle value on the scale whenever they experience a conflict between the conclusion based on the content of the arguments and the conclusion based on the pragmatic meaning of the instruction words. Two different age groups were examined: a group of 8-to-10-year-olds (Experiment 1) and a group of 10-to-12-year-olds (Experiment 2). The results of both age groups were similar. The percentages appropriate answers seemed to be higher for the older children in most cases but this difference was only significant for the 'SeSe *so*'– and the 'IrSe *so*' sentences. More importantly, we made predictions based on Table 1 which allow us to gain insight into children's sensitivity to the conventional implicature from *but* on the one hand and the content of the arguments on the other hand.

The scale data included evidence that children are aware of the pragmatic meaning of *but*, *so* and *nevertheless*. We found that the children in both experiments provided the neutral answer about one third of the time for the 'IrSe *nevertheless*'- and the 'SeIr *so*' sentences and even a considerable amount of appropriate answers. This is evidence that children are sensitive to the pragmatic meaning of *but*, *so* and *nevertheless.* If they would have been exclusively sensitive to the content then we would have expected almost 100% inappropriate answers. The answers on the 'SeIr *nevertheless*' sentences also indicate that children are sensitive to the pragmatic meaning of *but*. Both the content of the arguments and the use of *nevertheless* elicit the conclusion from *p*. However, *but* elicits the conclusion from *q* and this conclusion has to be cancelled in order to reach the appropriate conclusion. The fact that 30% (Experiment 1) and 21% (Experiment 2) neutral answers were given, suggests that sensitivity to the implicature from *but* causes doubt.

Apart from evidence showing that children are sensitive to the pragmatic meaning of the instruction words, the scale data also indicated that the content of the

arguments has a lot of influence on children's answers. When both arguments are sensible, none of the two arguments outweighs the other. When these sentences are combined with *nevertheless*, then *but* and *nevertheless* lead to opposite conclusions. In both experiments, one third of the answers were inappropriate which indicates that the pragmatic meaning of *nevertheless* is not that easy to grasp. The neutral answers (25% in Experiment 1 and 32% in Experiment 2) are interpreted as evidence that the children notice the conflict between the conclusion based on *but* and the conclusion based on *nevertheless*. As a consequence, the results from the 'SeSe *nevertheless*' sentences seem to suggest that children generally understand the pragmatic meaning of *but* and *nevertheless* but this understanding is far from perfect. When two sensible arguments are combined with *so*, both *but* and *so* elicit the conclusion from *q*. This means that the neutral (31% in both experiments) as well as the inappropriate answers (28% in Experiment 1 and 16% in Experiment 2) are evidence that the content of the arguments is very important for children and sometimes outweighs the answer based on the pragmatic meaning of *but* and *so*.

In contrast to the results showing that children are sensitive to the pragmatic meaning of *but*, *so* and *nevertheless* as well as to the content of the arguments, the results of the 'IrSe *so*' sentences pose a bigger problem to interpret. Both the content and the instruction words elicit the conclusion from *q* so we would have expected almost exclusively appropriate answers. The fact that 'only' 55% (Experiment 1) and 70% (Experiment 2) appropriate answers were provided suggests that children's performance on these 'p but q' sentences is far from optimal and the use of the middle answer on the scale might rather express a certain general incomprehension. So, this latter finding causes us to interpret our scale data cautiously. Further research with the use of scales on conventional implicatures seems necessary. It might be useful to apply this scale format on adults. This would allow us to compare their responses with children's responses. Consequently we could get a clearer view on how to interpret these results.

References

Anscombre, J., & Ducrot, O. (1977). Deux 'mais' en français? *Lingua, 43*, 23–40.

Berk, L.E. (2010). *Development through the lifespan* (5th ed.). Boston: Allyn and Bacon.

Braine, M., & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology, 31*(1), 46-70.

Grice, P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M. C. (1997). *Algemene Nederlandse Spraakkunst*. Groningen/Leuven: Wolters Noordhoff.

Janssens, L., & Schaeken, W. (2013). 'But' how do we reason with it: An experimental investigation of the implicature stemming from 'but'. *Journal of Pragmatics, 57*, 194-209.

Janssens, L., Drooghmans, S., & Schaeken W. (in press). But: do age and working memory influence conventional implicature processing? *Journal of Child Language.*

Katsos, N., & Bishop, D.V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67-81.

Noveck, I.A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition, 78*(2), 165-188.

Van Belle, W. (2003). *Zwijgen is niet altijd toestemmen. De rol van inferenties bij het interpreteren en argumenteren.* Leuven: Uitgeverij Acco.

Van Belle, W., & Devroy, G. (1992). *Tegenstellende en toegevende connectoren. Een argumentatieve beschrijving* (Preprint 143, voorlopige publicatie). Katholieke Universiteit Leuven, Faculteit Letteren, Departement Linguïstiek.

Chapter 6

‘But’ load doesn’t matter: The automaticity of processing conventional implicatures

Abstract

The objective of this study was to reach a clear conclusion regarding the possible cognitive costs underlying the processing of conventional implicatures. Adult participants were asked to perform a conventional implicature task in which they had to indicate the appropriate conclusions (introduced by *so* and *nevertheless*) stemming from ‘p but q’ sentences. Additionally, while performing the implicature task, working memory was burdened with a secondary task in four conditions ranging from no working memory load to a high load. The results showed that working memory load didn’t influence participants’ performance on the conventional implicature task. This finding suggests that, contrary to conversational implicatures, working memory is not involved in inferring conventional implicatures. Even though performance on the implicature task was far from perfect, the inferences from *but*, *so* and *nevertheless* seem to be automatically triggered.

Keywords: conventional implicatures; *but*; processing costs; working memory load

Janssens, L., & Schaeken, W. (submitted). ‘But’ load doesn’t matter: The automaticity of processing conventional implicatures. *Manuscript submitted for publication at Quarterly Journal of Experimental Psychology.*

Introduction

When people communicate with each other, they tend to follow a set of guidelines – called maxims by Grice (1989) - that ensure that the conveyed message is easily understood by all interlocutors. In order to avoid misinterpretation, people are expected to express themselves as informative as necessary. For example:

(1) Frank drank two beers.

The utterance in (1) can be interpreted as '*Frank drank exactly two beers*'. If Frank would have drunk six beers, (1) would still be logically true. However, the hearer can assume that the interpretation of exactly two beers holds because the speaker wants his utterance to be optimally understood by the hearer. The inference from (1) that Frank drank exactly two beers is an example of an implicature. This refers to an inference that consists of attributing to a speaker an implicit meaning that goes beyond the explicit linguistic meaning of an utterance (Noveck, 2001). A whole range of experimental research has focused on one specific type of implicature, namely conversational implicatures. The example described in (1) is a scalar implicature which is a subtype of conversational implicatures. A substantial part of the experimental scalar implicature research has focused on the cognitive processes underlying these inferences. Indirect evidence suggesting that deriving scalar implicatures is cognitively effortful can be found in developmental research. Noveck (2001), among others, found that children are more logical than adults. Children treated word such as *might* and *some* logically (*might* is compatible with *must* and *some* is compatible with *all*) whereas adults treated these words rather pragmatically (as exclusive to respectively *must* and *all*). Because children's cognitive capacities aren't fully developed yet, this was considered as indirect evidence that working memory capacities are involved in deriving scalar implicatures. More direct evidence was presented by De Neys and Schaeken (2007) who burdened participants' working memory capacity by providing them with a secondary task during performance of the scalar implicature task. When working memory was burdened, pragmatic inferences dropped with 10%. Likewise, Bott and Noveck (2004, Experiment 4) manipulated the availability of cognitive resources by varying response time. Participants had either three seconds or were limited to only 900ms to provide their answer. The number of pragmatic answers was 16% lower in this latter condition than in the former condition, indicating that pragmatic inferences require processing costs.

The implicatures stemming from the connector *but*, described in this paper, are part of a different category of implicatures, namely conventional implicatures. Horn (2004) defines this concept as follows:

> "Unlike an entailment or logical presupposition, this type of inference is irrelevant to the truth conditions of the proposition. This inference is not cancellable without contradiction, but it is detachable, in the sense that the same truth-conditional content is expressible in a way that removes (detaches) the inference. Such detachable, but non-cancellable aspects of meaning that are neither part of, nor calculable from, 'what is said' are conventional implicatures." (Horn, 2004, pp. 4)

The materials used in this experiment consist of 'p but q' sentences in which *but* is used as an indirect distancing contrastive connector. In this type of sentences, *but* is used as an 'argumentative *but*' (Van Belle, 2003). The first phrase of the 'p but q' argumentation is accepted as true, but the use of *but* causes the second phrase to have more weight and consequently to deny the inference from *p*. For example:

(2) I would like to go for a walk, but it's really cold outside.

The inference from the *p*-argument is that the speaker in (2) will go outside for a walk and the inference from the *q*-argument is that the speaker will not go outside. Because *but* is used, the *q*-argument gets more weight and consequently the conclusion follows that the speaker will not go outside for a walk. The conclusion from a 'p but q' sentence can be introduced by words like *so* or *nevertheless*. The pragmatic meaning of *so* elicits the inference from *q* as the appropriate conclusion (e.g., I would like to go for a walk, but it's really cold outside. So I won't go outside) and therefore strengthens the inference from *but*. In contrast, the pragmatic meaning of *nevertheless* cancels the inference from *but* and elicits the inference from *p* as the appropriate conclusion (e.g., I would like to go for a walk, but it's really cold outside. Nevertheless I will go outside). Janssens and Schaeken (2013) presented 63 adult participants with such 'p but q' sentences followed by either two *so*-conclusions or two *nevertheless*-conclusions and they were asked to indicate the appropriate conclusion. Janssens and Schaeken (2013) aimed to find out whether people truly understand the pragmatic meaning of these words and therefore choose the inference from *p* when the conclusion with *nevertheless* is asked and the

inference from *q* when the *so*-conclusion is asked. The *p*- and *q*-arguments were either both sensible or a combination of a sensible and an irrelevant argument. The conventional implicatures stemming from *but*, *so* and *nevertheless* should lead to a certain conclusion, irrespective of the (relevance of the) content of the arguments. The results showed that, although people clearly understood the pragmatic meaning of these words, the content of the arguments greatly influenced participants' answers. When a sensible argument was combined with an irrelevant argument, participants mostly based their conclusion on the sensible argument without taking into account the pragmatic meaning of *but*, *so* and *nevertheless*. Even when a combination of two sensible arguments was presented, performance was not perfect. This implies that the content of the arguments often prevails over the implicatures that could be drawn from the 'p but q' sentences. A different finding from Janssens and Schaeken (2013) was that *nevertheless* elicited more inappropriate answers (i.e. not in line with the conventional meaning of the instruction word) than *so*. They argued that this might be attributed to the fact that the pragmatic meaning of *nevertheless* doesn't actually evoke the inference from *p* as was expected. However, a different explanation of this finding seems more likely: In order to reach the appropriate *nevertheless*-conclusion from *p*, the implicature from *but* (i.e. the inference from *q*) has to be overruled which seems likely to be effortful. This finding, together with the general finding that drawing these implicatures doesn't happen flawlessly induces the possibility that processing these conventional implicatures is cognitively effortful and therefore requires cognitive capacity. Moreover, a closer look at the conventional implicature from *but* as an indirect distancing contrastive connector reveals that this implicature has certain properties of conversational implicatures. One specific feature that characterizes conversational implicatures but not conventional implicatures is that they are cancellable. However, there seem to be two implicatures coming from *but*. First, there are sentences in which *but* connects two parts and the use of *but* creates a contrast between the two parts. For example:

(3) She is blonde, but she is intelligent.

The use of *but* in (3) elicits the implicature that being blonde contrasts with being intelligent (at least in the speaker's view) although this contrast is not explicitly expressed.

Second, *but* can also be used in sentences in which the inference from the *p*- and *q*-argument already contrast each other and the implicature from *but* indicates that the second part of the argumentation (*q*) attains more weight. The former seems indeed to be a classical conventional implicature, i.e. a non-cancellable implicature. However, this is not true for the latter, which is the type that will be discussed in this paper. For example, by using *nevertheless,* the implicature from *but* is cancelled. *Nevertheless* denies the inference from *but* and guides the hearer or reader towards the inference from *p*. As a consequence, because they have certain features of conversational implicatures, the conventional implicatures discussed in this paper may not be purely conventional. This similarity with conversational implicatures is a different reason why the possibility arises that processing these conventional implicatures requires cognitive capacities. As mentioned earlier, the involvement of working memory capacities in scalar implicature processing has been shown directly (e.g., De Neys & Schaeken, 2007) and indirectly (e.g., Noveck, 2001) by comparing scalar implicature competence between adults and children.

Janssens, Drooghmans and Schaeken (in press) performed the same experiment with the exact same materials as Janssens and Schaeken (2013) but their participants were children aged 8 to 12. Additionally, working memory capacity was measured by using the Listening Span task (Daneman & Carpenter, 1980). Their results were similar to the adult results in Janssens and Schaeken (2013) but although a direct comparison between adults and children wasn't made, children's competence with these conventional implicatures seemed worse than adults' competence. Because children's working memory capacity isn't yet fully developed, this cautiously seems to be in line with the hypothesis that working memory is involved in processing conventional implicatures. However, no effect of the direct working memory measure on children's performance was found. This finding, in turn, suggests that working memory would not be involved in processing conventional implicatures.

In sum, it seems that there are good reasons to expect why working memory would be involved in conventional implicature processing. These arguments are threefold: (1) The experimental findings from Janssens and Schaeken (2013) that adults' competence with these implicatures is far from perfect. Especially processing *nevertheless* can be expected to require cognitive resources since it cancels the implicature from *but*. (2) The similarity between the conventional implicatures

discussed in this paper and conversational implicatures, which have been shown to require cognitive capacities (e.g., De Neys & Schaeken, 2007). (3) The findings from Janssens et al. (in press) that children seem to be less competent than adults in processing these conventional implicatures. As for conversational implicatures, this difference in competence can be explained by referring to children's limited cognitive capacity compared to adults.

In contrast, there are also good reasons to expect why working memory would not be involved in processing conventional implicatures. First, the finding that a measure of working memory capacity didn't influence conventional implicature processing in children (see Janssens et al., in press) is an experimental argument that suggests no involvement of working memory. In addition, not only this experimental evidence suggests no involvement of working memory capacity in conventional implicature processing, but this prediction can also be derived from Grice's theory. For example, Moeschler (2012) argued that "….This is a very important point in Grice's definition of a conversational implicature, because only conversational implicatures are supposed to be worked out. When an implicature is automatically triggered, through a reference to the meaning of a word, the implicature is conventional." (Moeschler, 2012, pp. 417).

In conclusion, it seems that there are both good reasons to expect why working memory would be involved in conventional implicature processing, as well as good reasons to expect why this would not be the case.

In this paper, we aim to answer the question whether working memory is involved in processing the conventional implicatures from *but*, *so* and *nevertheless*. We will try to reach a clear conclusion regarding the supposed processing cost of conventional implicature processing by building further on Janssens and Schaeken (2013). However, some adaptations are made in order to provide some sharper measures. First, since the content of the arguments seems to play an important role, we will use a more ecologically valid measure to study the effect of the content. Participants will be presented with weak and strong arguments instead of respectively irrelevant and sensible arguments as in Janssens and Schaeken (2013). Second, the effect of the instruction word *but* will be assessed by comparing performance on sentences including *but* with sentences in which the arguments are simply juxtaposed. Finally, we will manipulate working memory, not by directly measuring working memory capacity, but

by using the same paradigm as De Neys and Schaeken (2007) used in scalar implicature research. We will look at the effect of working memory load on conventional implicature competence by imposing a secondary task on participants that burdens working memory capacity.

Experiment

Method

Participants

A total of 210 undergraduate students from the University of Leuven (Belgium) with a mean age of 19.2 participated in our experiment. They were all native Dutch speakers and received course credit in exchange for participation.

Materials

Implicature Task. Every participant was presented with 16 short context stories, adopted from Janssens and Schaeken (2013). These stories, programmed in E-Prime 1.1, were presented on a computer and were followed half the time by two 'p but q' constructions and half the time by two 'p . q' constructions. For example, (translated from Dutch):

> *Mom and Ella are shopping. Ella sees a lovely teddy bear lying on the shelves. She asks Mom if she can have the teddy bear. Mom is not sure.*
>
> *Mom thinks: "Ella has been bad, but she lost her teddy bear."*
>
> or
>
> *Mom thinks: "Ella has been bad. She lost her teddy bear."*

After each argumentative construction (either 'p but q' or 'p . q'), the participants were asked to choose the appropriate *so*-conclusion (e.g., *'so Ella can have the teddy bear'* or *'so Ella cannot have the teddy bear'*) as well as the appropriate *nevertheless*-conclusion (*'nevertheless Ella can have the teddy bear'* or *'nevertheless Ella cannot have the teddy bear'*). Note that the participants were presented with a different pair of arguments from the same type before they had to judge the second conclusion type (e.g., *Mom thinks: "Ella already has a lot of teddy bears, but she's been very good lately."*). The

reason why we did this was to avoid that participants would notice that every appropriate *so*-conclusion from *q* implied the appropriate *nevertheless*-conclusion from *p*. Both the 16 stories and the *so*- and *nevertheless*-conclusions were presented in a random order. In contrast to Janssens and Schaeken (2013) we did not use irrelevant[1] arguments but we did make a distinction between weak and strong sensible arguments. In the example above, both the *p*- and the *q*-argument are strong sensible arguments. In the same context, an example of two weak sensible arguments is *Mom thinks: "I'm in a hurry, but it's a lovely teddy bear."* The reason for making the distinction between weak and strong sensible arguments instead of irrelevant and sensible arguments is that the contrast between a sensible and an irrelevant argument is too large. Since the irrelevant arguments have no relation to the context of the story (e.g., *Mom thinks: "Ella has been very good lately, but it's Saturday."*), the irrelevant arguments might sound so absurd that it actually makes sense to base your conclusion on the sensible arguments. By manipulating the strength of the arguments we can still look at the effect of content but with a more ecologically valid measure.

In order to choose plausible and good arguments for our constructions, we performed two pilot studies. In a first pilot study, 16 participants were instructed to read stories in which a person was always confronted with a 'dilemma' (e.g., a girl received some chocolates and has to decide whether or not to eat chocolate). We asked the participants to give both arguments why a person should or should not do something (e.g., 'being hungry' is an argument to eat chocolate; 'being allergic to chocolate' is an argument not to eat it). In a second pilot study we asked 16 different participants to rate the arguments that were generated in the first pilot study on a scale from 1 (very weak argument) to 7 (very strong argument). Based on these two pilot studies we created our experimental set. For both the constructions separated by a 'period' and the *but* constructions, there were four possible combinations of arguments: strong-strong, strong-weak, weak-strong and weak-weak. Moreover, we also took into account the axiological value of the arguments. Anscombre and Ducrot (1977) introduced this term to describe the argumentative orientation of arguments. The argumentative orientation can be positive or negative. A negative argument (e.g., Ella has been bad) is oriented towards a negative conclusion (she cannot have the teddy bear), whereas a positive argument (e.g., Ella lost her teddy bear) is oriented towards a

[1] Note that 'irrelevant' is labeled as 'absurd' in the original Janssens and Schaeken (2013) study.

positive conclusion (she can have the teddy bear). This led to a 2x2x2x4 design (two connectors: *but* or 'period' x two conclusion types: *so* or *nevertheless* x two axiological value combinations: negative-positive or positive-negative x four argument combinations: weak-weak, weak-strong, strong-weak and strong-strong).

Working memory Load task. We manipulated working memory load in order to determine whether the number of pragmatic responses would be lower when working memory is burdened. For our working memory manipulation, we used a secondary task based on the Double Task Paradigm used in De Neys and Schaeken (2007). We created four load conditions. In the low load condition, participants were presented with a 3x3 matrix with three dots that were always horizontally or vertically positioned. A matrix of this kind was displayed before every one of the 16 stories and participants had to remember the position of the dots in order to reproduce them in an empty matrix at the end of each implicature item. The moderate load condition was similar, but the dot pattern was more complex to remember. In this condition, participants were presented with a 3x3 matrix with four randomly positioned dots. In the high load condition, there were six randomly positioned dots in a 4x4 matrix. Finally, as a control, there was a no load condition in which the participants were not presented with matrices but were simply asked to perform the implicature task.

Procedure

The participants individually performed the task in groups of 50 students at the same time. All participants were presented with the 16 stories, followed by two questions about the appropriate conclusion. This means that every participant answered one item of every sentence type. Meanwhile the participants performed the working memory load task. The whole task lasted approximately 12 minutes per participant.

Results

First, we calculated the average number of correct reproduced dots in every load condition. Next, we removed all participants whose performance was less than 2 standard deviations below the average. We did this in order to exclude participants who didn't even try to perform well on the working memory task. In order to find a reliable effect of working memory load, every participant should take the working memory load task seriously. We removed two participants from the low load condition (n = 60), three

from the moderate load condition (n = 53) and one from the high load condition (n = 32). This left us with a total data set of 204 participants.

For our analyses a generalized linear mixed model with a logit link function was used (see e.g., Baayen, Davidson, & Bates, 2008; Bates, Maechler, & Bolker, 2011; or Jaeger, 2008). The model fitting procedure was implemented in R using the lmer() function from the lme4 package. We increased model complexity until the best model fit was reached. Model fit was assessed through the Bayesian Information Criterion (BIC).

The final model includes a three-way interaction between axiological value combination, argument combination and conclusion type and a two-way interaction between conclusion type and connector. We did not find an effect of working memory load: there were no significant differences between the load conditions in the mean accuracy scores. Table 1 displays a summary of the final model in which the intercept is compared with all other variables. However, t-tests were performed to further analyze the interactions in the model. Figure 1 displays the two-way interaction between connector and conclusion type. When the connector *but* separates the two arguments the mean accuracy score is significantly higher for *so*-conclusions (M = 0.64, SD = 0.48) than for *nevertheless*-conclusions (M = 0.51, SD = 0.5): t (3262) = 7.28, p <.001. However, when the two arguments are separated by a 'period', the mean accuracy scores don't differ significantly between *so*- and *nevertheless*-conclusions (*so*: M = 0.56, SD = 0.5; *nevertheless*: M = 0.54, SD = 0.5; t (3262) = 0.70, p = .48).

Table 1 *Parameter estimates for the final model*

Predictor	**Estimate**	**Standard Error**	***Z***	***p***
Intercept	0.20	0.11	1.75	.080
period	0.42	0.08	5.32	<.001***
nevertheless	0.04	0.15	0.24	.810
SW	0.74	0.15	5.0	<.001***
WS	1.41	0.16	8.81	<.001***
WW	0.90	0.15	5.99	<.001***
pos-neg	1.10	0.15	7.19	<.001***
period x nevertheless	0.56	0.11	5.17	<.001***
nevertheless x SW	0.93	0.21	4.48	<.001***
nevertheless x WS	2.69	0.22	12.14	<.001***
nevertheless x WW	1.45	0.21	6.98	<.001***
nevertheless x pos-neg	1.17	0.21	5.58	<.001***
SW x pos-neg	0.88	0.21	4.11	<.001***
WS x pos-neg	0.66	0.24	2.73	.006**
WW x pos-neg	1.96	0.21	9.15	<.001***
nevertheless x SW x pos-neg	1.88	0.30	6.19	<.001***
nevertheless x WS x pos-neg	1.57	0.32	4.89	<.001***
nevertheless x WW x pos-neg	2.85	0.30	9.63	<.001***

S=strong, W=weak
pos=positive, neg=negative
*p<.05, **p<.01, ***p<.001

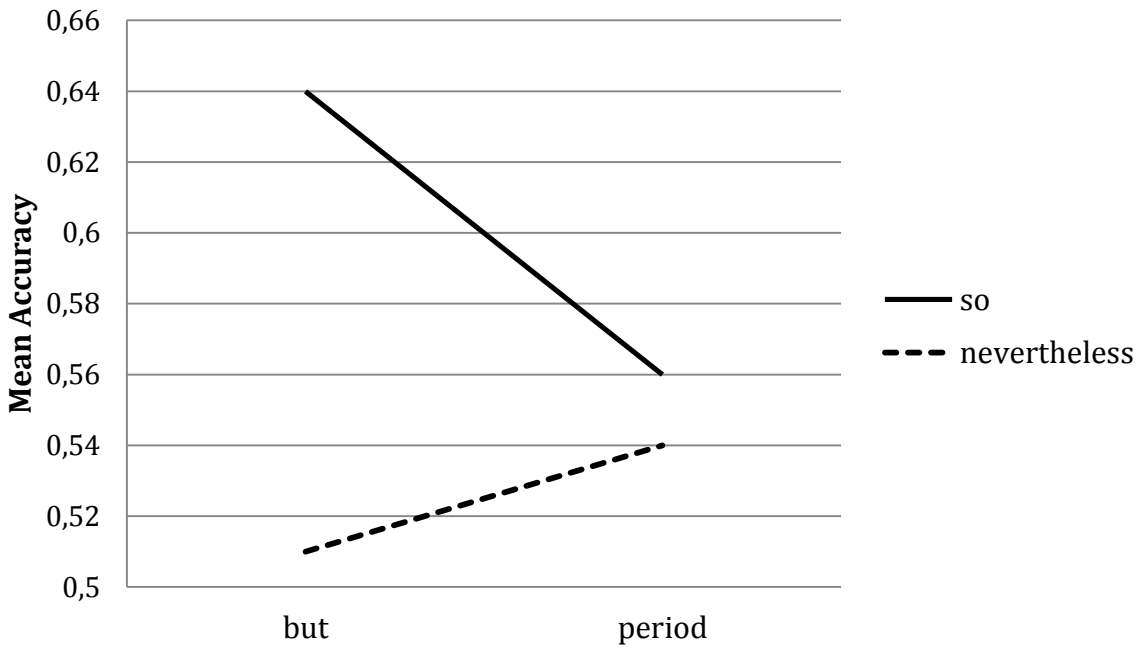


Figure 1. Interaction between connector and conclusion type.

Concerning the three-way interaction, Figure 2a, 2b, 2c and 2d display the interactions between conclusion type and axiological value combination for each of the different levels of argument combination. When a weak *p*-argument is combined with a strong *q*-argument, the axiological value combination 'positive-negative' leads to more accurate answers than 'negative-positive' for both *so* and *nevertheless* (*so*/neg-pos: *M* = 0.79, *SD* = 0.41; *so*/pos-neg: *M* = 0.85, *SD* = 0.36; *t* (814) = -2.28, *p* = .023) (*nevertheless*/neg-pos: *M* = 0.27, *SD* = 0.45; *nevertheless*/pos-neg: *M* = 0.45, *SD* = 0.50; *t* (814) = -5.41, *p* <.001). We find basically the same results for the combination of a strong *p*-argument with a weak *q*-argument with the exception that the difference for the *so*-conclusion is not significant (*so*/neg-pos: *M* = 0.33, *SD* = 0.47; *so*/pos-neg: *M* = 0.38, *SD* = 0.49; *t* (814) = -1.39, *p* = .17) (*nevertheless*/neg-pos: *M* = 0.60, *SD* = 0.49; *nevertheless*/pos-neg: *M* = 0.78, *SD* = 0.41; *t* (814) = -5.86, *p* <.001). When two arguments of the same strength are presented we see reversed patterns for strong-strong and weak-weak. In both these cases, it depends on the conclusion type whether 'positive-negative' or 'negative-positive leads to more accurate answers. When both arguments are weak, the axiological value combination 'negative-positive' leads to more accurate answers than 'positive-negative' for *so*-conclusions, but to less accurate answers for *nevertheless*-conclusions (*so*/neg-pos: *M* = 0.70, *SD* = 0.46; *so*/pos-neg: *M* =

0.51, *SD* = 0.50; *t* (814) = 5.68, *p* <.001) (*nevertheless*/neg-pos: *M* = 0.42, *SD* = 0.49; *nevertheless*/pos-neg: *M* = 0.62, *SD* = 0.49; *t* (814) = -5.56, *p* <.001). When both the *p*- and *q*-argument are strong arguments we find the reversed pattern, with the exception that the difference for the *nevertheless*-conclusions is not significant (*so*/neg-pos: *M* = 0.50, *SD* = 0.50; *so*/pos-neg: *M* = 0.74, *SD* = 0.44; *t* (814) = -7.27, *p* <.001) (*nevertheless*/neg-pos: *M* = 0.55, *SD* = 0.50; *nevertheless*/pos-neg: *M* = 0.54, *SD* = 0.50; *t* (814) = 0.49, *p* = .62).

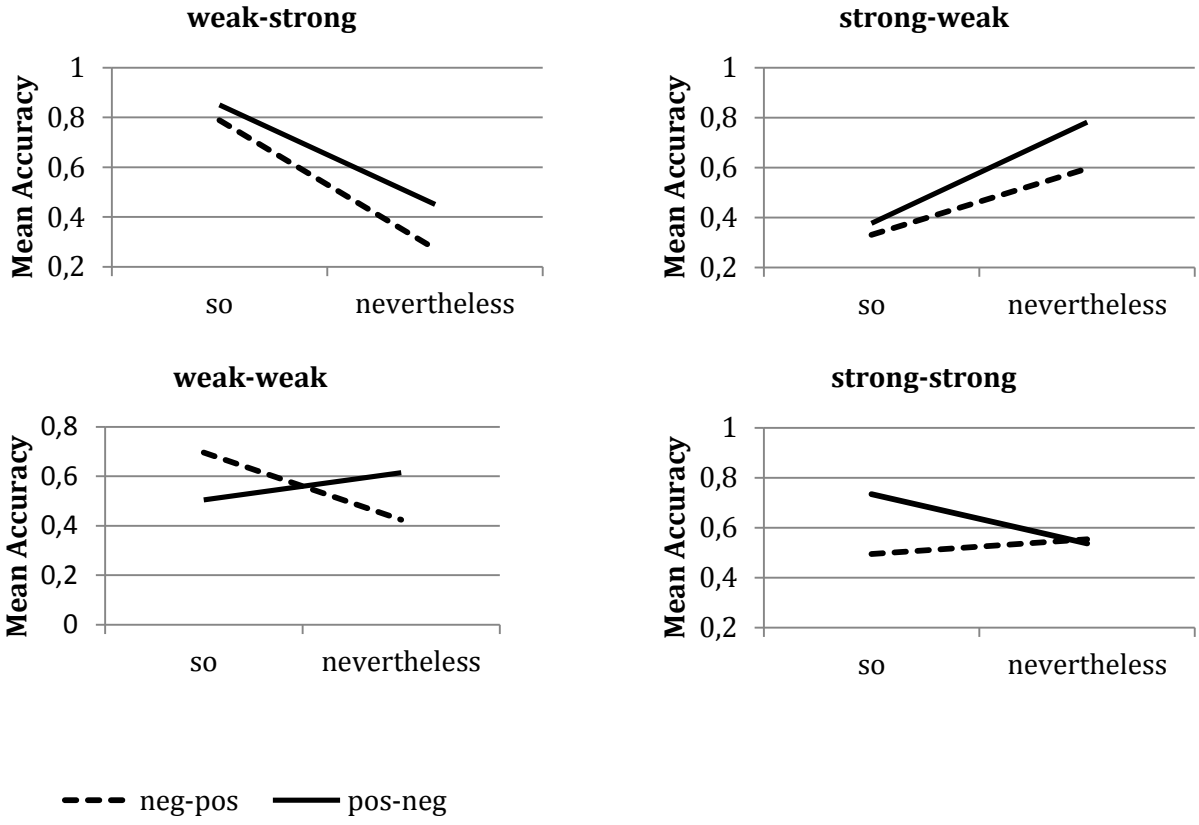


Figure 2a, b, c, d. Interactions between conclusion type and axiological value combination for each argument combination (a) weak-strong, (b) strong-weak, (c) weak-weak, (d) strong-strong.

General Discussion

Our manipulations enabled us to investigate the effect of several different variables. Compared to Janssens and Schaeken (2013) we optimized our manipulations: First, we made a comparison between the connectors *but* and 'period' so we would be able to explicitly look at the effect of *but*. Second, we made a distinction between weak and strong arguments instead of irrelevant and sensible arguments which provides a more reliable measure. Additionally, we had four working memory load conditions in order to explore whether a higher burden on working memory capacity would significantly decrease the number of appropriate answers.

We expected there to be an effect of connector, conclusion type, argument combination and working memory load. Our best fitting model for the data was a model including a two-way interaction between conclusion type and connector and a three-way interaction between conclusion type, argument combination and axiological value combination. This means that working memory load didn't affect the number of appropriate answers. This contrasts with findings in scalar implicature research that working memory is involved in processing these implicatures (e.g., Bott & Noveck, 2004; De Neys & Schaeken, 2007; Noveck, 2001). However, this finding is line with the results in Janssens et al. (in press) who measured working memory capacity in children and found no relation with their performance on the conventional implicature task. As a result, it seems that processing conventional implicatures happens differently from processing conversational implicatures. If working memory is not implied in inferring conventional implicatures, this suggests that processing these implicatures happens effortlessly. Moreover, the lack of a significant working memory effect also suggests that the poor performance on *nevertheless* sentences cannot be explained by the supposed cognitive costs involved in cancelling the implicature from *but* in favor of the implicature from *nevertheless*.

The two-way interaction depicted in Figure 1 provides evidence that the conclusion from *so* leads to more appropriate answers than the conclusion from *nevertheless*, at least when *but* separates the *p*- and *q*-argument. This means that the specific meaning of *but* is clearly understood and contributes to the understanding of *so* and *nevertheless*. The inference from *but* directs the reader towards the conclusion from the *q*-argument and the use of *so* following a *but* sentence confirms and strengthens this

conclusion. However, *nevertheless* requires the reader to overrule the inference from *but* in favor of the conclusion from *p*.

When a 'period' separates the *p*- and *q*-argument, there is no indication which of the two arguments has more weight and therefore what conclusion is the appropriate one. Consequently, there's no significant difference in the number of appropriate answers between *so*-conclusions and *nevertheless*-conclusions.

The three-way interaction is more difficult to interpret. We hadn't anticipated finding an effect of axiological value combination. As in Janssens and Schaeken (2013) – where this variable had no effect - axiological value combination was added as a control variable. Based on Figures 2a, 2b, 2c and 2d, we suggest a possible interpretation of this three-way interaction. Figure 2c depicts the situation in which two weak arguments are presented. It can be argued that this is not an obvious situation. Compared to weak-strong and strong-weak, none of the two arguments stands out over the other. Compared to strong-strong this construction only contains weak arguments and it may be less clear which inference stems from these weak arguments. Figure 2c shows that in this weak-weak situation, people make more correct inferences from a positive argument than from a negative argument for both *so* and *nevertheless.* This can be deduced from the fact that the axiological value combination 'negative-positive' leads to more appropriate *so*-conclusions than 'positive-negative' and the opposite applies for *nevertheless.* Since appropriate *so*-conclusions are inferred from the *q*-argument and *nevertheless*-conclusions from the *p*-argument, this means that positive arguments facilitate the appropriate conclusion in weak-weak situations. The same seems to hold for other less obvious situations with different argument combinations. We found that overall, *nevertheless*-conclusions elicited more inappropriate conclusions than *so*-conclusions and for these *nevertheless*-conclusions a positive argument seems to facilitate the appropriate conclusion compared to a negative argument as well. This can be seen in Figure 2a (weak-strong) and 2b (strong-weak). However, this does not hold for the *nevertheless*-conclusions when *p* and *q* are both strong arguments. In those sentences, there was no significant difference between 'positive-negative' and 'negative-positive'.

When we look at the *so*-conclusions, a reversed pattern seems to emerge. For both the weak-strong (Figure 2a) and the strong-strong (Figure 2d) situations, the axiological value combination 'positive-negative' leads to significantly more appropriate

so-conclusions than 'negative-positive' which implies that a negative argument facilitates the appropriate conclusion in these situations. This difference between 'positive-negative' and 'negative-positive' is not significant for the *so*-conclusions in the strong-weak situations (Figure 2b). This can be explained by the fact that this is the least obvious *so*-conclusion to make since it requires the reader to ignore a strong argument in favor of a weak argument.

Since the effect of axiological value combination was unexpected, the proposed interpretation remains suggestive. The reason why a negative argument would facilitate the appropriate conclusion in some situations whereas a positive argument facilitates the appropriate conclusion in other situations remains unclear. Further research is necessary to clear out why this difference between a positive and a negative argument occurred.

In spite of this unexpected three-way interaction and the lack of a load effect, in general most of the findings from our experiment are in line with our expectations. This experiment enabled us to provide evidence that people generally understand the conventional meaning of *but*. They understand that *but* causes the *q*-argument to attain more weight than the *p*-argument. Concerning the different conclusion types we found that more inappropriate answers are given when participants have to infer the appropriate *nevertheless*-conclusion than when they have to infer the appropriate *so*-conclusion. *So* confirms the inference stemming from *but* whereas *nevertheless* requires the reader to overrule the inference from *but* and return to the *p*-argument. A third expected finding is that the content of the arguments plays a very important role. Even though participants understand the conventional meaning of *but*, whenever a strong argument is combined with a weak argument they mostly base their conclusion on the strong argument and consequently ignore the conventional meaning of *but* (and *so* and *nevertheless*). This finding is identical to the effect of argument combination found in Janssens and Schaeken (2013) who combined sensible arguments with irrelevant arguments. Combining strong arguments with weak arguments provides a more reliable and ecologically valid measure. Despite of using this better measure, the same effect is found: even sensible arguments can get overruled simply because they're weak and measured against a stronger argument.

In sum, this paper aimed to provide a clear conclusion regarding the possible cognitive processes underlying the derivation of the conventional implicatures from *but*,

so and *nevertheless*. Our results showed that participants under a high working memory load didn't perform significantly different from participants under a low working memory load or participants whose working memory wasn't burdened at all. This is evidence that processing these implicatures happens automatically, without involvement of working memory. This finding is consistent with the finding in Janssens et al. (in press) who found no effect of a direct working memory measure on conventional implicature processing in children. This finding also provides evidence in favor of Grice's theory that claims that conversational implicatures have to be worked out but conventional implicatures are automatically triggered (see Moeschler, 2012). Overall, we can conclude from this study that conventional and conversational implicatures aren't derived the same way.

Concerning all other findings in this experiment, most were consistent with previous findings in Janssens and Schaeken (2013). The only exception is the incidental finding of the role played by axiological value combination. More research is necessary to provide a clear explanation for this unexpected effect.

References

Anscombre, J., & Ducrot, O. (1977). Deux 'mais' en français? *Lingua, 43*, 23–40. Doi: 10.1016/0024-3841(77)90046-8

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412. doi:10.1016/j.jml.2007.12.005

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes, R package version 0.999375-42 http://www.R-project.org/

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437-457. doi:10.1016/j.jml.2004.05.006

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466. Doi: *10.1016/S0022-5371*(*80*)*90312-6*

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology, 54*, 128-133. doi:10.1027/1618-3169.54.2.128

Grice, H. P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Horn, L. R. (2004). Implicature. In L. R. Horn, & G. Ward (Eds.), *The handbook of pragmatics* (pp. 3-28). Oxford: Blackwell.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*, 434-446. doi:10.1016/j.jml.2007.11.007

Janssens, L., & Schaeken, W. (2013). 'But' how do we reason with it: An experimental investigation of the implicature stemming from 'but'. *Journal of Pragmatics, 57,* 194-209. doi:10.1016/j.pragma.2013.08.021

Janssens, L., Drooghmans, S., & Schaeken W. (in press). But: do age and working memory influence conventional implicature processing? *Journal of Child Language.*

Moeschler, J. (2012). Conversational and conventional implicatures. In H-J. Schmid (Ed.), *Cognitive pragmatics* (pp. 405-436). Berlin: De Gruyter.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78*, 165–188. doi:10.1016/S0010-0277(00)00114-1

Van Belle, W. (2003). *Zwijgen is niet altijd toestemmen. De rol van inferenties bij het interpreteren en argumenteren.* Leuven: Uitgeverij Acco.

Chapter 7

It's maybe somewhat difficult *but* I understand it!

Abstract

Most studies of implicatures focused on conversational implicatures. This study however, examined the conventional implicature induced by *but*. According to the literature, one can assume that the second argument in a 'p but q' construction is the argument with most weight. This is, however, never experimentally tested with a direct distancing contrastive *but*. We presented participants with stories which ended with a direct distancing *but* construction, in which one of the arguments expressed a feeling of understanding towards the behavior of the main character in the story. The results indicated that indeed the *q*-argument has most weight. There was, however, also an effect of the specific content of the stories. These results are discussed in light of the hypotheses generated on the basis of previous research with an indirect distancing contrastive *but*, but also in the light of the effect of content of the stories in conventional implicature research and specific task characteristics.

***Keywords**:* conventional implicature; *but*; scale; content

Janssens, L., Delombaerde, K., & Schaeken, W. (submitted). It's maybe somewhat difficult *but* I understand it! *Manuscript submitted for publication in Psychologica Belgica*.

Introduction

As Clark and Schober (1992) formulated: *"It is a common misperception that language use has primarily to do with words and what they mean. It does not. It has primarily to do with people and what they mean. It is essentially about speakers' intention"* (Clark & Schober, 1992, p. 15). What we want to convey in daily communication is to a large extent not explicitly expressed. Instead, people in conversation make use of facial expressions, gesticulation, and the (assumed) intentions of the speaker to make their interactions successful. Grice (1975) is one of the founding fathers of pragmatics and provided us with a theoretical framework to discuss this issue. Starting point was the general principle of cooperation, which Grice (1975) formulates as follows: *"Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged"* (Grice, 1975, p.46). The cooperative principle can be divided into four maxims (Grice, 1989). The Maxim of Quantity entails that utterances should always be informative. The Maxim of Quality refers to the fact that this information should always be truthful. The Maxim of Relation in its turn expresses that utterances should always be relevant to the goals of the conversation and finally, the Maxim of Manner requires interlocutors to phrase everything appropriately. According to Grice (1989), whenever people follow these maxims, the result is an efficient exchange of information. However, these maxims are not exhaustive.

By means of the cooperative principle and the maxims, Grice (1989) describes the inference process, the retrieval of a speaker's meaning. This brings us to the term 'implicature' which refers to the inferred intentions of what a speaker didn't explicitly say. In his work, he made a distinction between two categories of implicatures, i.e. conversational implicatures on the one hand and conventional implicatures on the other hand. The idea of implicatures became quickly popular both in theoretical and experimental pragmatics. However, experimental research on implicatures has almost completely concentrated on (generalized) conversational implicatures. Conversational implicatures are dependent on the context and rely on the maxims of conversation. One has to be even more precise: most experimental research focused on the subcategory of scalar implicatures (e.g., De Neys & Schaeken, 2007; Dieussaert, Verkerk, Gillard, & Schaeken, 2011; Doran, Baker, McNabb, Larson, & Ward, 2009; Noveck, 2001; Papafragou & Musolino, 2003). Horn (1972) developed this concept. Horn-scales involve

a set of alternative expressions of the same grammatical category, but with a difference in semantic informativeness (for a recent theoretical review, see e.g., Geurts, 2010). The most well-known examples of such scales, ordered from strong to weak are <*all, most, some*> and <*and, or*>. Underlying these scales is the assumption that the use of a semantically weaker term implies that the stronger one does not hold. This assumption is based on the maxim of quantity of Grice (1989), that is, we want to be as informative as required, but also not more informative than necessary. When someone says '*I read some of the chapters of the PhD dissertation*', this person usually implies that he did not read all of the chapters. In other words, the logical meaning of *some* (at least one) receives a more precise interpretation in which *all* is excluded.

Scalar implicatures are examples of generalized conversational implicatures, which are assumed to be context-independent. They arise in such a broad range of contexts that it will not seem to depend on the details of the context too much. In contrast, there are also particularized conversational implicatures, which were assumed to be less systematic and always clearly context-dependent. An example of such a particularized implicature is the situation in which one wonders where the hamburger is and the grandmother suddenly says: "*Well, the dog is looking very happy*". In such a situation, we will derive the implicature that the grandmother thinks that the dog ate the hamburger. The derivation from 'looking happy' to 'did eat the hamburger' can only be made in this very specific context. Although some theorists treat the distinction between generalized and particularized implicatures as important (see e.g., Levinson, 1989), others argue against this distinction because generalized implicatures are still context-dependent, although they are more context-independent than particularized implicatures (see e.g., Hirschberg, 1985; Sperber & Wilson, 1995). Still others find the distinction between generalized and particularized implicatures theoretically inert (e.g., Geurts, 2010; Neale, 1992).

In the current study we will not focus on conversational implicatures, but on the seldom investigated conventional implicature. Conventional implicatures are independent of the cooperative principle: A statement always carries its conventional implicature, but this implicature is not part of the semantic meaning of the terms. On the basis of the different definitions found in the literature, Horn (2004) came up with a summarizing definition of this concept:

> "Unlike an entailment or logical presupposition, this type of inference is irrelevant to the truth conditions of the proposition. This inference is not cancellable without contradiction, but it is detachable, in the sense that the same truth-conditional content is expressible in a way that removes (detaches) the inference. Such detachable, but non-cancellable aspects of meaning that are neither part of, nor calculable from, 'what is said' are conventional implicatures." (Horn, 2004, p.4)

Huang (2006) defines a conventional implicature as "*a non-truth-conditional inference which is not deductive in any general, natural way from the saying of what is said, but arises solely because of the conventional features attached to particular lexical items and/or linguistic constructions".* (Huang, 2006, p.54)

More specifically, in the present study we investigated the implicature induced by the conjunction *but*. The word *but* (translated from Dutch *maar*) is the most commonly used connector to express a contrastive concessive relation (Van Belle & Devroy, 1992). This 'p but q' relation is a particular type of contrast in which one part of the utterance (*p*) is a concession and the other, contrastive part of the utterance (*q*) denies the inference that could be made based on *p* (Van Belle, 2003). In the Algemene Nederlandse Spraakkunst (ANS; *General Dutch Grammar*), three types of the connector *but* are distinguished (Haeseryn, Romijn, Geerts, de Rooij, & van den Toorn, 1997).

First, *but* can be used in a dividing contrast, in which *but* can be replaced by *and*. Replacing *and* with *but* emphasizes the contrastive nature of the connection, but not the other way around. Compare for instance 'he is rich and he is friendly' with 'he is rich but he is friendly'.

Second, *but* can be used in a replacing contrast. In such a construction, the first part of the sentence is a negation and the second part replaces the first part by expressing what's true (e.g., not cats but dogs are my favorite pets).

Finally, in a distancing contrast, *but* connects two parts of a complex speech act and the second part is disassociated from the first part, without denying what is being expressed in the first part (Van Belle & Devroy, 1992). In this type of 'p but q' construction, it is recognized by the speaker that *p* is true (Van Belle, 2003). However, the conventional meaning of *but* entails that the possible inference derived from *p* is cancelled. There are two possible ways in which this cancellation can be manifested (Moeschler & de Spengler, 1982).

On the one hand, *q* can directly cancel the inference from *p* because it contains a conclusion that contradicts the inference from *p* (*p (p → r), but q (q = not-r) (so not-r)*). For example: "The water is cold (→ will not swim), but I will swim in it". The connector *but*, used in a direct distancing contrast, is labelled as a 'concluding *but*'. In a direct concession, *p* and *q* are always connected by their content: *p* expresses a cause or a good reason for *r* and '*but q*' expresses the conclusion that contrasts with the expected conclusion *r* from *p*. It's because of this connection that *but* can be replaced or strengthened by a conjunctional adverb such as *nevertheless* (e.g., the water is cold, *nevertheless* I will swim in it).

On the other hand, *q* can indirectly cancel the inference from *p* because *q* contains an argument that can be considered as stronger or more relevant than *p*. (*p (p → r), but q (q → not-r) (so not-r)*). E.g.: "The water is cold (→ will not swim), but I like swimming (→ will swim). (So I will swim)". Note that it is the conventional meaning of *but* that causes the argument from *q* to overrule the argument from *p*. When the two arguments trade places, the opposite conclusion follows because the *q*-argument always outweighs the *p*-argument. E.g.: "*I like swimming, but the water is cold. So I will not swim"*. The connector *but*, used in an indirect distancing contrast, is labelled as an 'argumentative *but*'. Anscombre and Ducrot (1977) postulated three claims concerning this type of 'p but q' utterance:

1. By uttering 'p but q', the speaker always expresses some kind of acceptance of *p*.
2. *q* is always the argument with most weight and the 'p but q' construction must be viewed as a defense of not-*r*.
3. 'p but q' constructions are always aimed at cancelling a particular conclusion *r*.

When the word *so* (translated from Dutch *dus*), follows a 'p but q' utterance, it introduces and confirms the expected conclusion from *q* (e.g., "The water is cold, but I like swimming. So I will swim."). In contrast, whenever the word *nevertheless* (translated from Dutch *toch*) follows a 'p but q' utterance, it reverses the argumentative orientation. Consequently, *nevertheless* redirects the reader towards the conclusion stemming from *p* (e.g., "The water is cold, but I like swimming. Nevertheless, I will not swim.").

Given the high frequency of the connector *but*, it is surprising that there is almost no empirical research about this connector. One of the exceptions is Janssens and Schaeken (2013). They investigated the indirect distancing contrast use of *but*. In their research, they presented adult participants with short stories. Each of these stories ended with a 'p but q' sentence, which was followed by two possible conclusions. The participants were instructed to indicate the appropriate conclusion. These were either two *so*-conclusions ('*so* conclusion from *p*' and '*so* conclusion from *q*') or two *nevertheless*-conclusions ('*nevertheless* conclusion from *p*' and '*nevertheless* conclusion from *q*'). The appropriate pragmatic conclusion following *so* is the conclusion inferred from *q* and the appropriate conclusion following *nevertheless* is the conclusion from *p* (see Van Belle, 2003). The experiments showed that adults indeed understand the pragmatic meaning of *but*: *so*-conclusions primarily followed the *q*-argument and the *nevertheless*-conclusions followed the *p*-argument, although the preference was smaller. A plausible reason for the latter effect is the difficulty of *nevertheless*: one has to negate the negation of the expectation based on the *p*-argument

Interestingly, the content of the arguments also had an effect. In their first experiment, Janssens and Schaeken (2013) presented not only sensible arguments, but also irrelevant arguments. In the swimming example above, both arguments are sensible in a context in which a person doubts whether or not he will jump in the water. In this same context, uttering *"I like swimming, but I have a brother"* clearly contains an irrelevant *q*-argument. These irrelevant arguments were included to examine whether the pragmatic meaning of *but* is understood and used irrespective of the content of the arguments. This was not the case. It was observed that in those cases where an irrelevant argument was combined with a sensible argument, the participants had a clear preference for the conclusion from the sensible argument. This was true for both the *so*-conclusions and the *nevertheless*-conclusions.

In a second experiment, Janssens and Schaeken (2013) asked participants to justify their answers. Whenever participants did not provide the appropriate conclusion, they mostly motivated their answer by mentioning the content of the arguments.

The present research builds on the work of Janssens and Schaeken (2013), but there were three innovations.

First, instead of investigating the indirect distancing contrast use of *but*, in the current experiment the direct distancing contrast use of *but* will be examined. It is

important that one does not simply assumes that in different uses of *but* similar patterns will be observed. Indeed, experiments with scalar implicatures are for the most part confined to just two scalar expressions, *some* and *or*. Van Tiel (2014) showed convincingly that different scalars give rise to different amounts of pragmatic interpretations (see also Doran et al., 2009; Doran, Ward, Larson, McNabb, & Baker, 2012). Therefore, comparing different uses of *but* and checking whether people seem to have a general understanding of the pragmatic meaning of *but* and the different weighting for the different uses of *but,* seems fruitful.

Second, in the current study, one argument of the 'p but q' construction always represents a relevant argument in daily life, which even might have repercussions for consoling talks. That is, we used as one of the arguments an expression of understanding. An example of such a sentence is:

> *I understand that after many attempts you lost the hope for reconciliation, but a good communication between the two of you is important for the entire company.*

In half of the problems, the '*I understand*' argument was the *p*-argument. For the other half of the problems, this argument was expressed in the second part of the sentence (*q*). To control for the real effect of *but*, half of the problems contained the connector *but*, whereas for the other half of the problems, the two arguments were separated by a 'period'. For example:

> *I understand that after many attempts you lost the hope for reconciliation. A good communication between the two of you is important for the entire company.*

Third, the dependent variable was a different one from previous research. Instead of evaluating conclusions, participants were asked to express on a scale whether or not they expected that the person in the story would feel understood or not.

Our hypothesis was that there would be a significant interaction between order (whether *p* or *q* is the 'understanding argument') and connector (*but* or period). We expected the effect of order to be only there when the connector *but* is used instead of the period. Only in that case the *q*-argument has more weight, leading to higher feelings of being understood when the *q*-argument is the 'understanding argument'.

Experiment

Method

Participants

A total of 192 adults participated in the experiment. They were all psychology students at the University of Leuven and participated as part of a course requirement. All participants were native Dutch speakers.

Design

The experiment had a 2x2x2 design, whereby all independent variables were manipulated between participants. First, the connector was either *but* or a period. Second, the proposition in which the feeling of understanding was expressed was either the *p*-argument or the *q*-argument. Third, to control for possible content effects, we developed two different stories (one about a company, one about an exam). The dependent variable was a rating of the feeling of being understood of the main character.

Material and Procedure

All materials were in Dutch. Each of the stories started with a description of a very difficult situation. The company story (story A) goes as follows:

> Joke had a violent fight with her colleague. The close collaboration between them is important for a good functioning of the business. A misunderstanding that arose a few days ago, has escalated. Joke is convinced that her colleague made a mistake and does not want to concede. Her colleague is blaming Joke. Joke has repeatedly tried to talk about this, but this never led to a success. As a consequence, being in the same room leads inevitably to an angry passage of words. Therefore, Joke decided to not say a word to her colleague. Joke is very stubborn and determined to keep silent for the rest of their working collaboration.

After this introduction, the story continues with the introduction of the crucial manipulation:

> She talks about the situation with a different colleague. The colleague tells her: *"I understand that after many attempts you lost the hope for reconciliation, but a good communication between the two of you is important for the entire company."*

A quarter of the participants in the company condition received this story; for another quarter, the order of the two arguments was reversed; another quarter received the arguments in the same order, but instead of using the connector *but*, the sentences were now simply separated by a period; finally, another quarter received the two arguments in the opposite order, separated by a period. The university story (Story B) had the same four versions. An example of the crucial sentence in the university story is:

> *Carrying on with your study is important for your chances for a job later on, but I understand that you want to stop your study after such a dishonest act.*

Each participant received only one story. The participants were tested in five different groups, in which the different versions were distributed randomly. The participants were asked to imagine how the main character in the story would feel after the last sentence. They had to indicate this on a seven-point scale, going from 'feels totally misunderstood' to 'feels totally understood'.

Results and Discussion

Table 1 presents the average scores on the seven-point scale of the feeling of being understood for the different conditions. We performed an ANOVA, which resulted in two significant main effects and two significant interaction effects.

First, we observed a significant main effect of order. That is, when the expression of understanding is the *q*-argument, the feeling of being understood is higher than when it is the *p*-argument (4.01 vs. 3.16; F (1,188) = 19.37, $p < .05$, *partiële* $\eta^2 = .09$).

Second, there is, as expected, no significant main effect of the type of connector (*but*: 3.51; period: 3.66; F (1,188) = 0.48, $p > .05$, *partiële* $\eta^2 = .001$). This is because the variable order causes the scores on the scale for *but* to be in balance. The same is expected for the sentences separated by a period since none of the arguments is expected to outweigh the other. However, there is a significant interaction between order and connector *(*F (1,188) = 8.66, $p < .05$, *partiële* $\eta^2 = .04$). The effect of order is only there when the connector *but* is used instead of the period.

Third, to complicate things a little bit, there is a significant main effect of the variable story (story A vs story B: 3.95 vs 3.25; F (1,188) = 13.59, $p < .05$, *partiële* $\eta^2 = .07$) and an interaction between the variables story and order (F (1,188) = 7.25, $p < .05$, *partiële* $\eta^2 = .04$) indicating that the expected effect of order was only there in Story A

(4.17 vs 3.25). For Story B, the effect was in the expected direction, but not significant (3.40 vs 3.08).

Table 1 *Mean feeling of understanding score in the conditions with 'I understand' in the p- or the q-argument and with a period or but as a connector between the arguments*

Sentence type	Period	*But*
p - I understand	3.78	4.26
I understand - *q*	3.53	2.80

General Discussion

The present study contributes to the very recent experimental research into the area of conventional implicatures, and more precisely in the understanding of *but*. From the results of the present experiment, we can conclude that with a direct distancing contrast use of *but*, the *q*-argument indeed has a greater weight than the *p*-argument: Ratings of the expected feeling of being understood by the main character were clearly higher when the expression of understanding was in the *q*-argument instead of the *p*-argument. Importantly, this finding was only true when the two arguments were connected by *but*. When a period was used to connect the two arguments, there was no significant difference. The greater weight of the *q*-argument seems even higher than in the experiments of Janssens and Schaeken (2013) in which stories with an indirect distancing contrast use of *but* were presented. This might indicate that the claims of Anscrombre and Ducrot (1977) and Van Belle (2003) about the indirect distancing contrast use of *but* are not only true for the direct distancing contrast use as well, but even in a stronger way. However, we have to be careful with this conclusion because of two important problems or shortcomings of the current study.

First, there is the effect of content which was found in the current experiment. The expected effect was only significant in story A, the company story. In story B, the exam story, the trend was in the same direction, but the effect was not significant. Such an effect of content is not very surprising. Janssens and Schaeken (2013) also observed a strong content effect on the understanding of 'p but q' sentences. Therefore, one could argue that the observed effect was not due to the direct distancing contrast use of *but*,

but due to the effect of content. The fact that for Story B the effect was, although non-significant, in the same direction as for Story A, strengthens our belief in the observed significant effect. Nevertheless, we admit that further research is definitely needed.

Second, there is a difference in the task we used. Janssens and Schaeken (2013) asked to evaluate which of the given conclusions was the most appropriate. In the current experiment, participants were asked to express on a scale whether or not they expected that the person in the story would feel understood or not. Katsos and Bishop (2011) compared two different tasks in which participants had to evaluate scalar implicatures. In one experiment, they instructed their participants to judge on a binary scale (right vs wrong) how well a fictional character described certain situations. They observed what is typically observed in such binary judgment tasks when an underinformative sentence was presented, that is, a sentence in which *some* is used while *all* is also the case (e.g., using the sentence 'The crocodile played with some of the cars' while it was shown that the crocodile played with all the cars): Children do not penalize such a description as false whereas adults do. In a second experiment, they used a three-point scale with different sized strawberries. Now participants were instructed to reward a bad conclusion with the smallest strawberry, a conclusion that was neither completely bad nor good with the medium-sized strawberry, and a good conclusion with the biggest strawberry. As a result, children's performance did not differ anymore from adults': The underinformative sentences were judged by both groups with the middle value on the scale. This indicated that the use of the scale can reveal children's comprehension of scalar implicatures whereas a binary task conceals their competence. Although we did not use a ternary scale in this experiment, it is clear that it has more resemblances with a ternary scale than with a binary scale. Therefore, one could argue that it is the type of task that caused the effect and not specifically the direct distancing contrastive use of *but*. Further research has to confirm if it was the type of answer-scale that is a crucial factor.

Furthermore, in our task, participants did not have to evaluate whether or not an utterance was right or wrong (or something in between), but they had to imagine how the main character in the story would feel after the last sentence and express it on a seven-point scale, going from 'feels totally misunderstood' to 'feels totally understood'. Again, further research must clarify whether or not this dependent variable was crucial in finding the straightforward effects of *but*.

References

Anscombre, J-C., & Ducrot, O. (1977). Deux 'mais' en français? *Lingua, 43*, 23–40.

Clark, H.H., & Schober, M.F. (1992). Asking questions and influencing answers. In J.M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology, 54*, 128-133.

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology, 64*, 2352-2367.

Doran, R., Baker, R., McNabb, Y., Larson, M., & Ward, G. (2009). On the non-unified nature of scalar implicature: an empirical investigation. *International Review of Pragmatics, 1*, 1–38.

Doran, R., Ward, G., Larson, M., McNabb, Y., & Baker, R. (2012). A novel paradigm for distinguishing between what is said and what is implicated. *Language*, *88*(1), 124–154.

Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.

Grice, H.P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics, Vol. 3* (pp. 41-58). New-York: Academic Press.

Grice, H.P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M. C. (1997). *Algemene Nederlandse Spraakkunst*. Groningen/Leuven: Wolters Noordhoff.

Hirschberg, J. (1985). *A theory of scalar implicature*. Doctoral dissertation, University of Pennsylvania.

Horn, L.R. (1972). *On the semantic properties of logical operators in English*. Unpublished doctoral dissertation, University of California, Los Angeles (Distributed by Indiana University Linguistics Club, 1976).

Horn, L.R. (2004). Implicature. In L.R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics*. Oxford: Blackwell.

Huang, Y. (2006). *Pragmatics.* New York: Oxford University Press.

Janssens, L., & Schaeken, W. (2013). 'But' how do we reason with it: An experimental investigation of the implicature stemming from 'but'. *Journal of Pragmatics, 57*, 194-209.

Katsos, N., & Bishop, D.V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67-81.

Levinson, S.C. (1989). Review of Relevance: Communication and cognition. *Journal of Linguistics*, *25*, 455-472.

Moeschler, J., & De Spengler, N. (1982). La concession ou la réfutation interdite, approches argumentative et conversationnelle. *Cahiers de linguistique française, 4*, 7-37.

Neale, S. (1992). Paul Grice and the philosophy of language. *Linguistics and philosophy, 15*, 509–559.

Noveck, I.A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition, 78*, 165-188.

Papafragou, A., & Musolino, J. (2003) Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*, 253-282

Sperber, D., & Wilson, D. (1995). *Relevance (2nd edition)*. Oxford: Blackwell.

Van Belle, W. (2003). *Zwijgen is niet altijd toestemmen. De rol van inferenties bij het interpreteren en argumenteren.* Leuven: Uitgeverij Acco.

Van Belle, W., & Devroy, G. (1992). *Tegenstellende en toegevende connectoren. Een argumentatieve beschrijving* (Preprint 143, voorlopige publicatie). Katholieke Universiteit Leuven, Faculteit Letteren, Departement Linguïstiek.

van Tiel, B. (2014). *Quantity matters: Implicatures, typicality and truth.* Doctoral dissertation, Radboud Universiteit Nijmegen.

Chapter 8

Summary and Final discussion

In this dissertation, the two major research lines of my doctoral project were discussed. These two parts correspond with the two broad categories of implicatures, namely conversational and conventional implicatures. The conversational implicature study focused on the scalar implicature from *some* whereas the conventional implicature studies focused on *but*, combined with *so* and *nevertheless*. Both research topics were addressed from a developmental point of view. For both types of implicature the same aspects were investigated: working memory, age and task characteristics.

In the final chapter of this dissertation, we will first summarize the findings from each of the described studies. Next, in a final discussion, we will integrate the most important conclusions of this doctoral research within the broader theoretical framework of implicatures. We will successively discuss our conclusions with regard to what they revealed about the role of task characteristics, development and working memory.

Summary of the experimental findings

Chapter 2

Chapter 2 described a developmental study on the scalar implicature from *some*. Even though this domain has already been widely experimentally investigated, there were some aspects yet to uncover.

First, various studies have been carried out with adults and children in different languages such as Greek (e.g., Papafragou & Musolino, 2003), English (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Katsos & Bishop, 2011), French (Noveck, 2001, Experiment 3; Pouscoulous, Noveck, Politzer, & Bastide, 2007), Italian (Foppolo, Guasti, & Chierchia, 2004), German (Doitchinov, 2005), etc. In the Dutch language, there are scalar implicature data available on Dutch speaking adults, but not on children. It has been previously shown that scalar implicature performance is language-dependent. For example, Pouscoulous et al. (2007, Experiment 3) showed that even within a language,

there is a difference in processing the scalar implicature from French *certains* and *quelques* which both correspond to *some.* The results of our study in Chapter 2 showed that seven- and even five-year-old children are already highly pragmatic reasoners. This seems to indicate that Dutch is a very accessible language for the derivation of the implicature from *some*.

Second, the participant group in the study described in Chapter 2 also included three-year-olds. Such young children usually don't get tested within the domain of experimental implicature research.

Third, working memory involvement had been shown in adult studies or indirectly by the finding that children seem to be less pragmatic than adults, but working memory had never been directly measured in children.

Finally, the importance of task characteristics is another aspect that is already widely documented in previous literature (e.g., Chierchia et al., 2001; Katsos & Bishop, 2011). However, we compared a linguistic task with an action task and fine-tuned the manipulations of these task characteristics in order to become a valid and clear comparison. Additionally, we also manipulated the specific task content in Experiment 3.

The results of Experiment 1 in Chapter 2 revealed a reliable effect of age: five-year-olds provided significantly more pragmatic answers than three-year-olds in both a TVJT and an ABT. Three-year-olds clearly favored the logical meaning of *some.* This was not only clear from the answers on the underinformative items but also from the control sentences that could be answered logically, pragmatically or wrong. However, contrary to our expectations, no effect of task was found: the ABT didn't elicit significantly more pragmatic answers than the TVJT in Experiment 1.

In contrast, Experiment 2 did reveal a significant effect of task characteristics in a group of five-year-old children. As expected, the more difficult TVJT elicited more logical answers than the ABT.

A reliable effect of working memory could not be found in both Experiment 2 and Experiment 3 (seven-year-olds). Children with a higher working memory capacity did not provide significantly more pragmatic answers than low capacity children. Experiment 3 did show that not only the task itself influences the number of pragmatic answers, but also the specific task content. A TVJT which required cognitive world-knowledge led to significantly more logical answers than a TVJT with simple visual input.

Chapter 3

Chapter 3 described the first study on conventional implicatures in this dissertation. Adult participants were presented with short stories ending with a 'p but q' sentence. The *p*- and *q*-arguments were combinations of sensible and irrelevant arguments. The participants were asked to indicate the appropriate *so*-conclusion from two options or the appropriate *nevertheless*-conclusion (also from two options). The collected data showed that the conventional meaning of *but* in 'p but q' sentences is clearly understood, but the content of the arguments is a non-negligible contributing factor in interpreting these sentences. When one of the arguments is irrelevant, this argument practically always gets ignored in favor of the sensible argument, even if the conventional meaning of *but* required the participant to choose the answer from the irrelevant argument. The importance of the content of the arguments was confirmed in Experiment 2 in which participants were instructed to provide reasons for their answers. Participants often referred to the content of the arguments when they didn't infer the appropriate conclusion.

A second finding from Chapter 3 is that the meaning of *so* seems to be well understood, but the meaning of *nevertheless* seems much less clear.

Chapter 4

The main goal of Chapter 4 was to investigate whether children understand the conventional meaning of *but*, *so* and *nevertheless* and if so, to find out whether an effect of age can be detected. Moreover, to explore the possible involvement of working memory, a group of high span children was compared with a group of low span children. The findings from this study were very similar to the adult findings described in Chapter 3. Children seemed to generally understand the meaning of *but* and *so*, whereas they had more difficulties with *nevertheless*. Even though a direct comparison with adults' interpretations wasn't made, children's understanding of this implicature seemed to be worse than adults'. However, within the group of 8-to-12-year-old children, no effect of age was found. Older children didn't perform more accurately than younger children. Likewise, high capacity children's performance didn't differ significantly from low capacity children's performance, suggesting no involvement of working memory in conventional implicature processing.

Chapter 5

In Chapter 5 of this dissertation, a different response format was used to investigate children's comprehension of conventional implicatures. Inspired by Katsos and Bishop (2011), we expected a three-point scale to reveal children's competence on a more detailed level. Specifically, we expected sentences in which there was a conflict between the answer based on the content of the arguments and the answer based on the conventional meaning of the instruction words to evoke the middle answer on the scale, mirroring that the conflict is experienced. In contrast, we expected sentences in which there was no conflict to be represented as an extreme answer on the scale. Contrary to these expectations, the children used the middle answer on the scale an equal number of times for each of the different sentence types. This made it difficult to interpret these results correctly and rather it suggested that children are not as competent with these conventional implicatures as previously assumed.

Chapter 6

Chapter 6 focused on the role of working memory in conventional implicature processing again. Different from the study in Chapter 4, working memory was investigated in adults by means of a dual task paradigm: working memory capacity was not directly measured but burdened by a secondary task while performing the implicature task. Moreover, our manipulations were optimized. First, we compared 'p but q' sentences with 'p . q' sentences so we could truly find out whether there is an effect of *but*. Second, instead of using irrelevant and sensible arguments, we used weak and strong sensible arguments. These correspond more closely to everyday language use. These manipulations increased the ecological validity of our experiment. The findings of the study in Chapter 6 showed that there was, as expected, a significant difference between 'p . q' and 'p but q' sentences. This means that the conventional meaning of *but* indeed provides more weight to the *q*-argument. Moreover, the weak and strong arguments had the same effect as irrelevant and sensible arguments. Finally, concerning the working memory manipulation, our results did not reveal a significant effect of working memory, confirming and strengthening the conclusion from Chapter 4 that working memory is not involved in processing the conventional implicatures from *but*, *so* and *nevertheless*.

Chapter 7

In Chapter 7, the focus was on direct distancing contrasts instead of indirect distancing contrasts. In direct distancing contrasts, one of the arguments always expresses a conclusion that contrasts the inference from the other argument. Moreover, the 'p but q' sentences were of a very specific kind: there always was an argument that expressed understanding involved (e.g., *Carrying on with your study is important for your chances for a job later on, but I understand that you want to stop the study after such a dishonest act*). As in Chapter 6, we compared 'p but q' sentences with 'p . q' sentences. The results of this study revealed that, parallel with the indirect use of *but*, the *q*-argument indeed has a greater weight than the *p*-argument in the direct use as well. This finding applied when the two arguments were connected by *but,* but not when the arguments were separated by a period. However, importantly, the content of the sentences played a role as well. Two different contexts were used to embed the items in a short story and contrary to our expectations, only one of the stories yielded the expected effects.

Final discussion

In the final section of this dissertation, I will discuss the three aspects that were investigated in this doctoral research project in both conversational and conventional implicatures, namely task characteristics, development and working memory.

Task characteristics

Conversational implicatures

It has been shown in previous research that competence with scalar implicatures depends heavily on the kind of task that is used. For example, Grodner, Klein, Carbary and Tanenhaus (2010) found results that contrasted those found by Huang and Snedeker (2009) concerning the processing cost underlying the computation of the scalar implicature from *some*. Grodner et al. (2010) based their explanation of why these findings differed from each other on differences in the used stimuli. This shows that small differences in the task can have far reaching consequences for the results and therefore the conclusion. Katsos and Bishop (2011) on their part, have shown that using a scale instead of a binary response format reveals that adults and children are equally

aware of pragmatic anomalies, but that children are simply more tolerant towards pragmatic violations of underinformativeness.

One could say that experimental settings are not representative for every day communication. However, Bonnefon, Feeney and Villejoubert (2009) studied scalar inferences from *some* in very specific situations that correspond to real life events. In daily life, people can be confronted with face-threatening or face-boosting conditions. Brown and Levinson (1987) define *face* as a sense of positive identity and public self-esteem that all humans project and are motivated to support in social interactions (in Bonnefon et al., 2009, p. 250). In Experiment 1 of Bonnefon et al. (2009), participants were confronted with two different stories that ended with either a face threat (e.g., *some people hated your poem*) or a face boost (e.g., *some people loved your poem*). Participants were asked 'if it is possible from this utterance that everyone hated/loved your poem?' The results showed that *some* is rather typically (pragmatically) interpreted in the face-boosting condition as 'some but not all' whereas the face-threatening condition elicited significantly more 'yes' answers, indicating that the logical 'some and possibly all' interpretation is significantly more likely when there is a face threat. This finding indicates that not only task characteristics, but also specific communication contexts – politeness considerations in this case - influence how likely it is that a scalar inference will be derived.

Apart from specific task characteristics or specific communication contexts, the type of scalar also has an influence on the likeliness of deriving a scalar implicature. Van Tiel (2014) addressed in his doctoral dissertation the matter of scalar diversity. He argued that the majority of scalar implicature research has focused on the two scalar expressions *some* and *or*. According to van Tiel (2014), several classes of scalar expressions have been overlooked in experimental research. For example, adjectives (e.g., <*warm, hot*>), adverbs (e.g., <*sometimes, always*>) and nouns (e.g., <*mammal, dog*>) received very little attention and even within the classes that did get investigated, the variety of scalar expressions is limited. The reason for this very limited scope of investigated scalar implicatures is the underlying assumption that these scalar expressions are representative for scalar implicatures in general. However, this assumption of scalar uniformity has been tested by Doran, Baker, McNabb, Larson and Ward (2009) who found that there is a significant variability in the rates at which scalar terms of different lexical classes elicit pragmatic inferences (Doran et al., 2009, in van

Tiel, 2014). Van Tiel (2014) also showed that the rates at which scalar expressions give rise to pragmatic inferences are very diverse and that the scale <*all,* some> is an extreme case. In sum, the type of scalar is another aspect linked to task characteristics that determines how easily a scalar implicature will be derived.

In this dissertation, we looked into the difference between a linguistic task and an action task. This difference had already been investigated in Pouscoulous et al. (2007) but because of their multiple manipulations they weren't able to disentangle the effect of task from other possible effects. Our results confirmed that an action task elicits more pragmatic answers in children than a linguistic task which is assumed to be more cognitively effortful. This finding in its turn seems to provide support for the hypothesis that scalar implicature processing is cognitively effortful which will be discussed further on in the final discussion of working memory involvement.

A second finding, that had not been tested in any previous literature, is that two identical tasks (i.e. evaluating the truth value of statements) lead to different results because their specific content differs. Simple, visually available stimuli lead to a higher number of pragmatic inferences than non-visual stimuli that are based on world-knowledge.

Together, these results confirm and strengthen the conclusion that task characteristics always have to be taken into account when drawing conclusions on conversational (scalar) implicature processing. It's necessary to think in which way a different task, different stimuli or a different type of scalar might lead to different results. What's likely to happen in one context isn't necessarily true for a different context. There is no general rule about when scalar implicatures are (not) derived. It not only depends on the specific scalar term, but also on the type of task, the specific task content or even the specific communication context in daily life communication (see Bonnefon et al., 2009).

Conventional implicatures

As mentioned earlier in this dissertation, little experimental literature on conventional implicatures is available. Consequently, compared to conversational implicatures, little is known about how task characteristics influence the derivation of conventional implicatures. However, since Katsos and Bishop (2011) revealed that scales allow a better insight in scalar implicatures compared to binary judgment tasks, we argued that the same might be true for conventional implicatures. We noticed from our first

conventional implicature studies that the content of the *p*- and *q*-arguments often interfered with drawing the appropriate conventional conclusion in 'p but q' sentences. We hypothesized that using scales might reveal that children experience a conflict between the content of *p* and *q* and the conventional meaning of *but*, *so* and *nevertheless*. Chapter 5 of this dissertation presented results on conventional implicature processing in children by using a scale instead of a binary choice format. In contrast to the scalar implicature study in Katsos and Bishop (2011), the use of a scale didn't reveal that children indicated the middle answer when they experienced conflict between the conventional meaning of the sentence and the conclusion based on the specific content of the arguments. Instead, children indicated the middle answer of the scale an equal number of times for all sentence types, irrespective of whether these were 'conflict sentences' or not. We cautiously interpreted these results as evidence that this kind of implicature processing is actually very difficult for children. In order to get more clarity on this issue, future research should make a direct comparison between a scale task and a binary choice task. Moreover, the use of scales should also be tested in adults. Because their competence with these sentences seems to be better than children's competence, adult scale results might enable us to draw firm conclusions about the role of task characteristics in conventional implicature processing. Based on the current results, we are unable to draw such conclusions.

The results of all the conventional implicature studies in this dissertation do show the importance of the content of *p* and *q* in 'p but q' sentences. Of course, this is not the same as actual task characteristics but it does show that it is a factor that strongly influences how often the conventional implicature from *but* is drawn. Moreover, Chapter 7 presented results on *but* as a direct distancing contrast in which one of the arguments expressed a feeling of understanding. We found that, in this experiment too, the content of the stories influenced the derivation of conventional implicatures since only one of the two stories resulted in line with our expectations. Except from a different context/content, there were no differences between the two stories.

Development

Conversational implicatures

All the studies in this dissertation were approached from a developmental point of view. The developmental question that arises is: 'at what point in development are children

fully competent pragmatic reasoners?' This question is almost impossible to answer since many factors contribute to how easily (scalar) implicatures are derived. For example, Guasti et al. (2005) found that 87% of the Italian speaking seven-year-olds in their study accepted statements such as *some giraffes have long necks* whereas only 50% of the adults did. They used the same stimuli, translated in Italian, that were used in Noveck (2001) and that we also used in Chapter 2 of this dissertation. This finding would lead to the conclusion that the age of seven is still too early in development to be fully pragmatically competent. However, when these authors tested the same age group with a TVJT in which participants were asked to judge a puppet's description of a story acted out with toys and props, both the seven-year-olds and the adults rejected the underinformative *some* statements to the same extent. If we would only consider these results, then the conclusion would be that children aged seven are as pragmatically competent as adults. Consequently, the developmental question would be answered differently depending on which experiment is looked at. This shows, as we have discussed earlier in this dissertation, that task characteristics influence scalar implicature processing and thus that there is an interaction between answering the developmental question and task characteristics.

However, not only task characteristics, but also the specific scalar that is investigated influences how the developmental question could be answered. For example, Papafragou and Musolino (2003, Experiment 1) investigated Greek speaking five-year-olds. They used scalar terms from three different scales: <*all*, *some*>, <*three*, *two*> and <*finish*, *start*>. They found that a very low percentage of scalar implicatures was derived from *some* and *start* (respectively 12.5% and 10%) whereas the number of scalar inferences was a lot higher for the numeral scale (65%). This shows that not all scalar implicatures can be generalized. When trying to answer the question of when children are as pragmatically competent as adults, this conclusion will depend on the kind of scalar that was investigated. This too shows that it is impossible to put an exact age on when in development children are considered as pragmatically competent reasoners with scalar implicatures in general.

A third aspect that interacts with the developmental question is the specific language. It's difficult to answer the developmental question because it seems that different languages give rise to different results. Dutch speaking children's scalar implicature competence has rarely been investigated. Our results suggest that the Dutch

language is rather accessible to derive scalar implicatures, even in young children. In Experiment 3 of Chapter 2, seven-year-olds were shown to be almost exclusively pragmatic on underinformative sentences, at least on tasks with simple, visually available materials. We did not make a direct comparison with adults, but still, we can conclude from our experiments that even five-year-olds are already highly pragmatically competent. This age seems younger than results from previous studies in other languages (e.g., Noveck, 2001; Pouscoulous et al., 2007). The seven-year-olds in Experiment 3 of Chapter 2, who performed a world-knowledge TVJT that was translated to Dutch from the original stimuli in Noveck (2001), provided 69.4% pragmatic answers on the underinformative *some* statements. When we compare this with the original data from Noveck (2001), he found that seven-to-eight-year-old French speaking children only provided 11% pragmatic answers on the same *some* sentences. Ten-to-eleven year old children and adults respectively provided 15% and 59% pragmatic answers in the Noveck (2001) study. This shows that the seven-year-olds in our study were even more pragmatic than the adults in the Noveck (2001) study. This leads to the clear conclusion that the developmental question associated with scalar implicatures is language dependent. As Pouscoulous et al. (2007) have shown, even within a language there can be differences between different words that give access to scalar implicatures such as *certains* and *quelques* as French translations of *some*. In conclusion, when trying to answer the developmental question, the language involved is an important influential factor. In some languages such as Dutch, it seems that pragmatic competence is acquired earlier in development compared to other languages. Together with other factors such as task characteristics and the type of scalar, these variables influence the age found to be a critical point in development of pragmatic competence.

With regards to our own scalar implicature experiments in Chapter 2, we cannot provide an answer to the developmental question at what point in development children are fully competent pragmatic reasoners because of the important factors mentioned above. When we compared three-year-olds with five-year-olds on two different tasks, children as young as three years proved to be linguistically competent by performing well on the control sentences of both tasks. However, compared to five-year-olds, the three-year-olds provided a logical answer significantly more often on the underinformative sentences. We concluded from this experiment that the age of three is too young to show pragmatic competence, at least on the tasks used in our experiment

and for the Dutch language. Five-year-olds on the other hand were highly pragmatic, especially on the action task. Seven-year-olds proved to be even more pragmatically competent. This shows that there is a developmental trend, as we expected, but since the nature of the task and the task content clearly influenced the results, it is impossible to put an exact age on the moment when Dutch speaking children are fully pragmatically competent.

We conclude from all of this that the developmental question for conversational (scalar) implicatures interacts with factors such as language, task characteristics and type of scalar which makes it very hard to answer this question with one definite point in development. However, even though we cannot put an exact age on the development of scalar implicatures, there is abundant literature that shows that under the right experimental conditions, even very young children show a high level of pragmatic competence (e.g., Chierchia et al., 2001; Janssens & Schaeken, 2012; Katsos & Bishop, 2011; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004).

Conventional implicatures

Because experimental conventional implicature research is sparse, little is known about how the derivation of these implicatures develops. In contrast to scalar implicatures, no experiments are available stating that the comprehension of this specific conventional implicature from *but* is fully established at this or that age. As we discussed above for conversational implicatures, it is plausible that the same factors, such as language and task characteristics, influence the developmental question regarding conventional implicatures as well.

Our studies focused both on adults and on children between the ages of 8 and 12. We didn't directly compare children with adults, but it was clear that children's performance on this type of implicature was worse than adults'. However, within the child research, no effect of age was found: in a group of 8-to-12-year-olds, the older children didn't show more competence than the younger children. In fact, the results of Chapter 5, where children had to indicate their answers on a scale, elicited the suggestion that children might not be that competent with the conventional implicature from *but*, *so* and *nevertheless*.

The pragmatic delay compared to adults is likely to be explained by the great influence of the content of *p* and *q* in 'p but q' sentences. We already explained that the specific conventional implicature from *but* in our studies doesn't function as a typical

conventional term, indicating that there is a contrast between two arguments. Rather, it connects two already contrasting arguments and implies that the second argument (*q*) should have more weight than the first argument and therefore elicit the appropriate conclusion. We assume that children would have less problems with the typical conventional implicature from *but* (i.e. understanding that *but* implies a contrast) whereas they have more difficulty with *but* in our use because of the interference of the content of the arguments. Verbrugge (2007) performed a study in which she showed that children are used to think in terms of semantic content. Verbrugge (2007) investigated children (aged 6-7 and 11-12) and adults in two experiments. The participants were asked to complete 10 conditionals. They were presented with antecedents and were instructed to complete the sentence with a consequent of their choice (e.g., *If John is tired, then*…). Verbrugge (2007) was interested in how these conditionals would be completed spontaneously. The completion could result in either a content conditional (e.g., *…then he must go to bed early*) or an inferential conditional (e.g., *…then he has been working hard today*). She also included conditions in which participants were instructed to incorporate certain markers in their responses such as *may well* or *probably*. Verbrugge (2007) expected that adults would produce more inferential conditionals than children. A second prediction was that adults would use more inferential markers spontaneously compared to children. Finally, a third prediction was that participants would produce more inferential conditionals when they are asked to incorporate a particular inferential marker into their sentences. These three predictions were confirmed by the results. The youngest children produced the fewest number of inferential conditionals, the older children a few more. However, there was a significant gap with adults who produced significantly more inferential conditionals than children. This fits in with the literature (e.g., Bloom, Lahey, Hood, Lifter, & Fiess, 1980; Spooren, Tates, & Sanders, 1996) showing that children will generate more content relations spontaneously than inferential relations, claiming that this is because inferential relations are acquired later in development than content relations. According to Spooren et al. (1996), semantic relations are cognitively less complicated than pragmatic relations and that's why they are acquired earlier. These findings and claims may explain why the children in our experiments pay that much attention to the content and have trouble inhibiting this content in favor of the pragmatic meaning of the instruction words.

In sum, our binary scale results of Chapter4 showed evidence that children understand the conventional implicatures from *but*, *so* and *nevertheless* whereas the ternary scale results of Chapter 5 suggested that children's comprehension might not be that good. This makes it difficult to make any claims about children's competence and about when this competence is fully developed. We explained the difficulties children encounter with these conventional implicatures by referring to the influential role of the content of the arguments in 'p but q' sentences. Moreover, it is worth mentioning that even adults are influenced by this factor to a great extent. If we interpret the results of the scale study in Chapter 5 as indicating that children aged 8 to 12 are really too young to reason competently with such sentences, then it might be worth investigating a group of young adults, such as 15-year-olds in order to see if they already perform on the same level as adults. We can conclude for now that the results of this dissertation are too limited to draw clear and firm conclusions about the developmental aspect associated with conventional implicatures. Moreover, as we already mentioned above concerning conversational implicatures, it's likely that the same factors such as language, task characteristics and type of implicature influence any conclusion regarding the developmental question. For example, we assume that different types of conventional implicature develop differently from each other. Therefore, the findings from these experiments only apply to *but* used as an indirect distancing contrastive connector.

Working memory

Both the studies on conventional implicatures as well as the study on scalar implicatures, discussed in this dissertation regarding working memory involvement, reported no significant effects.

Conversational implicatures

Regarding scalar implicatures, a whole range of experimental literature has provided evidence in favor of the contextual theories that scalar inferences are not processed automatically (e.g., Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; De Neys & Schaeken, 2007; Dieussaert, Verkerk, Gillard, & Schaeken, 2011; Huang & Snedeker, 2009; Marty, Chemla, & Spector, 2013; Noveck & Posada, 2003). In contrast, other literature doesn't seem to find any processing costs for scalar implicatures. For example, Marty et al. (2013) found a working memory load effect associated with computing the scalar implicature from *some*, but an opposite working memory effect on quantity

implicatures. When working memory was burdened with a higher load, the number of pragmatic responses increased for quantity implicatures compared to the low load condition. Also, Feeney, Scrafton, Duckworth and Handley (2004) found that a measure of working memory capacity was significantly associated with providing the logical interpretation on infelicitous *some* statements. They argued that working memory is involved in inhibiting the pragmatic interpretation in favor of the logical one. Other evidence that suggests no role for working memory was provided by Grodner et al. (2010). They showed in a visual-world study that there was no delay associated with the pragmatic inference from *some* compared to other, non-scalar expressions. Likewise, in a visual world 'look-and-listen' study of Breheny, Ferguson and Katsos (2012) no difference was found between the time course of accessing scalar implicatures and the time course of accessing other contextual inferences.

The question whether or not processing costs are associated with inferring scalar implicatures is closely linked to the question of timing: relative to the literal interpretation, when does the pragmatic interpretation occur? One possibility is that the literal semantic meaning gets triggered automatically and the pragmatic meaning is only reached when this automatic interpretation gets cancelled. In Chapter 2 of this dissertation, we followed this line of reasoning and expected working memory to be involved in scalar implicature processing because the cancellation of the initial literal interpretation requires working memory. However, further on in this discussion of working memory, we will explain how we changed our view in the meantime based on an alternative explanation.

The results of our scalar implicature research in Chapter 2 didn't demonstrate a significant working memory effect. In contrast to most other scalar implicature research that investigated the role of working memory, our study directly measured working memory capacity instead of burdening working memory span with a secondary task. This difference may play a role. For example, Dieussaert et al. (2011) measured participants' working memory capacity and also burdened this capacity during the scalar implicature task. They found that a reliable working memory load effect (less pragmatic answers under cognitive load) could only be found for the low span group. This effect was absent in the moderate- and high span groups. However, looking more closely to their data, even though the load effect was present in the low span group, the low span group provided more pragmatic answers, over all conditions, than the other

two span groups. This means that if Dieussaert et al. (2011) would have just measured working memory capacity and looked for proof of working memory involvement by just comparing the span groups, they would not have found that lower working memory capacity is associated with fewer pragmatic responses. This is consistent with our findings in Chapter 2. This fact, together with the other ambiguous findings concerning working memory involvement in scalar implicature research, doesn't provide much clarity on this issue. It seems to suggest that, if working memory is involved, its contribution is rather limited. That's why we are unable to find a significant effect when comparing groups based on their cognitive capacity. It simply doesn't require that much cognitive resources under normal, full memory capacity access, circumstances. Memory load, on the other hand, strongly reduces available capacities and therefore affects scalar implicature processing, even though this effect is rather limited as well. A useful idea for future research might be to perform a meta-analysis, in which all the cognitive cost studies are summarized and compared. This will provide a clearer image of all the findings concerning working memory involvement in scalar implicature processing. This will also allow us to compare the results of studies that directly measured working memory and studies that burdened working memory with a secondary task.

Since the performed experiments described in Chapter 2, it has been brought to our attention that the relation between the timing of inferring a scalar implicature and its context dependency can be viewed independently from each other. According to Grodner et al. (2010) "On the one hand, it is logically possible that the inferential interpretation arises by default but that it takes time for this interpretation to become sufficiently activated to influence behavior. On the other hand, there are at least two context sensitive mechanisms of inferencing that predict the pragmatic meaning will be available immediately." (Grodner et al., 2010, p. 3). In other words, they argued that the integration of the inferential interpretation with relevant information in the context may require additional processing time. Consequently, this kind of reasoning does not exclude the possibility of working memory involvement. It's only suggested that working memory is not implied in cancelling the initial automatic literal interpretation. If one assumes that the logical and the pragmatic interpretation are complementary, then in order to reach the pragmatic interpretation it's not necessary for the logical interpretation to be cancelled, it simply needs to be enriched. This enrichment is what may require working memory resources. If we assume that working memory is involved

in the enrichment from the logical interpretation to the pragmatic interpretation, but only to a small extent, this might also explain why some studies report a quick automatic pragmatic inference (e.g., Grodner et al., 2010) whereas others do not (e.g., Huang & Snedeker, 2009). Only small differences in task design are sufficient to influence how quickly and easy this enrichment takes place. When a task is designed in such a way that it easily effectuates pragmatic enrichment, then working memory will not be required. This suggestive interpretation again stresses the importance of task characteristics, which we have already discussed in this final chapter of this dissertation.

As we mentioned above, in the final discussion of the task characteristics, it is also worth considering how specific communication contexts influence the cognitive costs associated with scalar implicature derivation. We already discussed the study of Bonnefon et al. (2009) in which they showed that implicit politeness rules influence the derivation of the scalar inference from *some*. In face-threatening contexts, the pragmatic interpretation of 'some but not all' is reached significantly less often than in face-boosting conditions. In a follow-up study, Bonnefon, De Neys and Feeney (2011) were interested in the processing costs of deriving scalar inferences in these face-threatening contexts. They replicated the finding that face-threatening contexts encourage the logical interpretation of 'some and possibly all'. Moreover, in line with other studies (e.g., De Neys & Schaeken, 2007) they found that face-boosting contexts elicited a drop in pragmatic responses under cognitive load (even though these pragmatic answers weren't associated with significantly longer response times). However, in face-threatening contexts, the opposite was found. Cognitive load increased the number of pragmatic inferences in these contexts and these inferences required shorter response times than logical interpretations. They concluded that 'politeness is hard to process'. This shows that specific communication contexts (in this case implicit politeness rules) influence cognitive costs associated with making the scalar inference.

Finally, in this discussion of working memory involvement in scalar implicature processing we cannot fail to mention research providing an alternative explanation than cognitive capacity for children's pragmatic incompetence. Barner, Brooks and Bale (2010) argued that children's difficulties to compute scalar implicatures are not associated with cognitive processing costs or other pragmatic limitations. According to them, children's difficulties can be explained by the fact that they lack knowledge of the relevant scalar alternatives to scalar terms like *some*. They tested this hypothesis in an

experiment with four-year-olds who were presented with pictures that depicted three out of three objects fit a description (e.g., three animals reading). They had to evaluate statements that relied on context-independent alternatives like *some* and *all* or on contextual alternatives (e.g., object labels such as dog and cat). These alternatives were sometimes preceded by *only* (e.g., *only some* of the animals; *only* the cat and the dog) and sometimes not. Barner et al. (2010) argued that, if children would understand *only* but fail to strengthen sentences with *only some*, this would be evidence that children are unable to generate scalar alternatives. Their results indeed showed that children didn't reject false statements containing context-independent scales, even when the word *only* was used (e.g., *only some* of the animals) whereas they did reject statements containing contextual alternatives (e.g., *only* the cat and the dog). These results were taken as evidence in favor of the hypothesis that children's difficulties with deriving scalar implicatures are caused by their inability to generate relevant alternatives for specific scales and are therefore unrelated to working memory capacity.

Conventional implicatures

Regarding our conventional implicature studies on *but*, practically no literature is available on the role of working memory in processing these implicatures. However, there were several reasons for us to expect that processing the implicatures from *but*, *so* and *nevertheless* would be cognitively effortful.

First, the 'p but q' sentences in our experiments consisted of arguments of which the inferred conclusions contradicted each other. For example:

(1) I really like chocolate, but it's almost dinner time.

In order to reach the appropriate conclusion from (1), one must infer from *p* that this is an argument that elicits the conclusion that the speaker will eat chocolate, whereas the inferred conclusion from *q* is that the speaker will not eat chocolate. Additionally, the conventional implicature from *but* implies that the second argument weighs more heavily so that the final conclusion is inferred from *q*. As a consequence, three inferences should be made in order for this final conclusion to be reached.

Theoretically, Iten (2005) and Blakemore (1987) claimed that *but* operates under a 'contradiction and elimination' principle. They both suggested that *but* contradicts and eliminates the inferred conclusion from *p*. Intuitively, assuming that all these steps are required in order to reach the appropriate conclusion from a 'p but q' sentence, it

seemed plausible to expect that processing these sentences might require working memory.

Second, as mentioned in Chapter 6, we noticed a certain similarity between the conventional implicature from *but* in our studies on the one hand and conversational implicatures on the other hand. The conventional meaning of *but* typically suggests a contrast between *p* and *q* whereas this contrast would not be suggested when *but* is replaced by *and* (e.g., she is cute, but she is smart). However, we investigated sentences in which there already is a contrast between the inferred conclusions from *p* and *q* and the conventional meaning of *but* implies that the conclusion from *q* is the appropriate conclusion. This type of inference can be cancelled by introducing the conclusion from *nevertheless* (e.g., I really like chocolate, but it's almost dinner time. Nevertheless I will eat chocolate). The characteristic of cancellability usually only applies to conversational implicatures (Geurts, 2010). That's why we wouldn't consider the 'p but q' sentences in our experiments as purely conventional. Because of its similarity to conversational implicatures, we expected the possible involvement of working memory in processing the conventional implicatures in our experiments.

The third reason we had to expect that processing these conventional implicatures might be cognitively effortful is derived from the use of *nevertheless.* It seemed very plausible that inferring the appropriate conclusion from *nevertheless* requires working memory since the implicature inferred from *but* has to be cancelled and the inference from *p* has to be activated again. Following the 'contradiction and elimination' view this would require an eliminated inference to be triggered again. It would be plausible that this process requires cognitive resources.

In contrast with our expectations, we found no evidence that working memory is involved in processing 'p but q' sentences as indirect distancing contrasts. From a theoretical point of view, this suggests that Hall's (2004) claim on how conventional implicatures from *but* are processed might be more accurate than the 'contradiction and elimination' perspective. According to Hall (2004), the *p*-argument isn't eliminated by *but*, it merely introduces a different argument that points in the opposite direction and that receives more weight because of the conventional meaning of *but*. This argumentation makes it more understandable why we didn't find that processing these implicatures is cognitively effortful. Another theoretical argument for not expecting a working memory effect is that according to Moeschler (2012), Grice's definition of

conversational and conventional implicatures implies that "….only conversational implicatures are supposed to be worked out. When an implicature is automatically triggered, through a reference to the meaning of a word, the implicature is conventional." (Moeschler, 2012, p. 417).

In sum, similar to our findings for scalar implicatures, we can conclude that working memory doesn't seem to be involved in processing *but* (together with *so* and *nevertheless*) in indirect distancing contrasts. The question remains why adults and especially children show such a poor performance on *nevertheless*-conclusions. If this poor performance cannot be explained by the cognitive effort it was supposed to require to draw these conclusions, then some other factor must be in play. It remains unclear what factor that might be. A different possibility is, however, that the conventional meaning of *toch* (translated in this dissertation as *nevertheless*) is not, as we expected, to cancel the implicature from *but* and to return to the inference from *p*. If that is true, then a simple experiment can clear out this possibility. We could present participants with sentences containing just one single argument in which *but* is absent. We would instruct our participants to provide the conclusion introduced by *nevertheless*. For example:

(2) I really like chocolate. Nevertheless……

When participants would complete this sentence in (2) with the conclusion that 'I will not eat chocolate', then it's clear that the true conventional meaning of *nevertheless* is indeed to cancel the inferred conclusion from the presented argument. In fact, our expectation is that people would indeed complete this kind of sentences as expected and therefore that the conventional meaning of *nevertheless* indeed cancels a certain belief or conclusion. This expectation is partly based on the similarity to German *doch*. Schmerse, Lieven and Tomasello (2014) found that, as they expected, the discourse particle *doch* involves making the inference that a prior belief state is re-assessed.

Furthermore, it is worth mentioning that we are very cautious and only draw this conclusion of no working memory involvement for the specific conventional implicature investigated in this dissertation. As Marty et al. (2013) have shown for scalar implicatures, working memory involvement depends on the type of scalar implicature (i.e. a significant working memory load effect for *some* but an opposite effect for quantity implicatures). The same might also hold for conventional implicatures. It's not because we did not find working memory to be associated with processing *but*, that this is true

for all other conventional implicatures as well. That's why we are planning to focus on different types of conventional implicatures, such as *already, also, barely, either, only, scarcely, still, too, yet* etc. in future research.

References

Barner, D., Brooks, N., Bale, A. (2010). Quantity implicature and access to scalar alternatives in language acquisition. *Proceedings of Semantics and Linguistic Theory, 20*, 525-543.

Blakemore, D. (1987). *Semantic Constraints on Relevance*. Oxford: Blackwell.

Bloom, L., Lahey, M., Hood, L., Lifter K., & Fiess, K. (1980). Complex sentences: acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language*, *7*, 235-261.

Bonnefon, J-F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition, 112*, 249-258.

Bonnefon, J-F., De Neys, W., & Feeney, A. (2011). Processing Scalar Inferences in Face-Threatening Contexts. *Proceedings of the Annual Conference of the Cognitive Science Society.* Cogsci. Boston, US, 3389-3394.

Bott, L., & Noveck, I.A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*(3), 437-457.

Breheny, R., Katsos, N., & Williams, J. (2006). Are Generalised Scalar Implicatures Generated by Default? *Cognition*, *100*(3), 434-463.

Breheny, R., Ferguson, H., & Katsos, N. (2012). Investigating the time-course of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, *28*(4), 443-467.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage.* Cambridge: Cambridge University Press (Original work published 1978).

Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. In A. H.–J. Do, L. Dominguez & A. Johansen (Eds.), *Proceedings of the 25th Boston University Conference on Language Development* (pp.157-168). Somerville, MA: Cascadilla Press.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology, 54*(2), 128-133.

Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for some: Further evidence that scalar implicatures are effortful. *The Quarterly Journal of Experimental Psychology, 64* (12), 2352-2367.

Doitchinov, S. (2005). Naturalistic and experimental data in language acquisition: The case of epistemic terms. In S. Kesper & M. Reis (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp. 123–144).

Doran, R., Baker, R., McNabb, Y., Larson, M., & Ward, G. (2009). On the non-unified nature of scalar implicature: an empirical investigation. *International Review of Pragmatics*, *1*(2), 1–38.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. (2004). The Story of *Some*: Everyday Pragmatic Inference by Children and Adults. *Canadian Journal of Experimental Psychology, 58*(2), 121-132.

Foppolo, F., Guasti, M.T., & Chierchia, G. (2004). *Pragmatic inferences in children's comprehension of scalar items.* Talk presented at Second Lisbon Meeting on Language Acquisition, Lisbon, 2004.

Geurts, B. (2010). *Quantity implicatures*. Cambridge: Cambridge University Press.

Grodner, D.J., Klein, N.M., Carbary, K.M., & Tanenhaus M.K. (2010): "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition, 116*(1), 42–55.

Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes, 20*(5), 667-696.

Hall, A. (2004). The meaning of but: A procedural reanalysis. *UCL Working Papers in Linguistics, 16*, 199-236.

Huang, Y.T., & Snedeker, J. (2009). Online Interpretation of Scalar Quantifiers: Insight into the Semantic-Pragmatics Interface. *Cognitive Psychology*, *58*, 376-415.

Iten, C. (2005). *Linguistic Meaning, Truth Conditions and Relevance. The Case of Concessives.* New York: Palgrave Macmillan.

Janssens, L., & Schaeken, W. (2012). The role of task characteristics in children's scalar implicature production. *Proceedings of the Annual Conference of the Cognitive Science Society*. Cogsci. Sapporo, Japan, 509-514.

Katsos, N., & Bishop, D.V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*, 67-81.

Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua, 133*, 152-163.

Moeschler, J. (2012). Conversational and conventional implicatures. In H-J. Schmid (Ed.), *Cognitive pragmatics* (pp. 405-436). Berlin: De Gruyter.

Noveck, I.A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition, 78*(2), 165-188.

Noveck, I.A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*(2), 203-210.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition, 86*(3), 253-282.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition, 12*(1), 71-82.

Pouscoulous, N., Noveck, I.A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition, 14*(4), 347-375.

Schmerse, D., Lieven, E., & Tomasello, M. (2014). Discourse particles and belief reasoning: The case of German *doch*. *Journal of Semantics, 31*, 115-133.

Spooren, W., Tates, K., & Sanders, T. (1996). Taalverwerving en de classificatie van coherentierelaties. *Nederlandse Taalkunde, 1*, 26-51.

van Tiel, B. (2014). *Quantity matters: Implicatures, typicality and truth.* Doctoral dissertation, Radboud Universiteit Nijmegen.

Verbrugge, S. (2007). *A psycholinguistic analysis of inferential conditional sentences*. Doctoral dissertation, Katholieke Universiteit Leuven.